FORM PTO-1390 (REV. 1-96) U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

ATTORNEY'S DOCKET NUMBER: 4239-50420

U.S. APPLICATION NO. (If known, see 37 CFR 1.5): 09/125635

INTERNATIONAL APPLICATION NO.	INTERNATIONAL FILING DATE	PRIORITY DATE CLAIMED
PCT/US98/12689	June 17, 1998	June 17, 1997

TITLE OF INVENTION: AIB1, A NOVEL STEROID RECEPTOR COACTIVATOR

APPLICANT(S) FOR DO/EO/US: Meltzer et al.

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. [X] This is a FIRST submission of items concerning a filing under 35 U.S.C. 371.
2. [ ] This is a SECOND or SUBSEQUENT submission of items concerning a filing under 35 U.S.C. 371.
3. [ ] This express request to begin national examination procedures (35 U.S.C. 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and PCT Articles 22 and 39(1).
4. [ ] A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.
5. [X] A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   a. [X] is transmitted herewith (required only if not transmitted by the International Bureau).
   b. [ ] has been transmitted by the International Bureau.
   c. [X] is not required, as the application was filed in the United States Receiving Office (RO/US).
6. [ ] A translation of the International Application into English (35 U.S.C. 371(c)(2)).
7. [X] Amendments to the claims of the International Aplication under PCT Article 19 (35 U.S.C. 371(c)(3))
   a. [ ] are transmitted herewith (required only if not transmitted by the International Bureau).
   b. [ ] have been transmitted by the International Bureau.
   c. [X] have not been made; however, the time limit for making such amendments has NOT expired.
   d. [ ] have not been made and will not be made.
8. [ ] A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371 (c)(3)).
9. [X] An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. [ ] A translation of the annexes of the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11. to 16. below concern document(s) or information included:

11. [X] An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. [ ] An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
13. [X] A FIRST preliminary amendment. *
    [ ] A SECOND or SUBSEQUENT preliminary amendment.
14. [ ] A substitute specification.
15. [ ] A change of power of attorney and/or address letter.
16. [X] Other items or information:
    8 Sheets of formal drawings

*Enter preliminary amendment before calculating filing fee.

U.S. APPLICATION NO. (If known, see 37 CFR 1.5)	INTERNATIONAL APPLICATION NO.	ATTORNEY'S DOCKET NUMBER
	PCT/US98/12689	4239-50420

17. ☒ The following fees are submitted:	CALCULATIONS PTO USE ONLY	
BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5)):		
Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO $1070.00		
International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO $930.00		
International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO $790.00		
International preliminary examination fee (37 CFR 1.482) paid to USPTO but all claims did not satisfy provisions of PCT Article 33(1)-(4) $720.00		
International preliminary examination fee (37 CFR 1.482) paid to USPTO and all claims satisfied provisions of PCT Article 33(1)-(4) $98.00		
ENTER APPROPRIATE BASIC FEE AMOUNT =	$ 790	00
Surcharge of $130.00 for furnishing the oath or declaration later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(e)).	$	

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE		
Total claims	54 - 20 =	34	x $22.00	$ 748	00
Independent claims	13 - 3 =	10	x $82.00	$ 820	00
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+ $270.00	$	
TOTAL OF ABOVE CALCULATIONS =				$ 2358	00
Reduction of 1/2 for filing by small entity, if applicable. A Small Entity Statement must also be filed (Note 37 CFR 1.9, 1.27, 1.28). +				$	
SUBTOTAL =				$ 2358	00
Processing fee of $130.00 for furnishing the English translation later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(f)).				$	
TOTAL NATIONAL FEE =				$ 2358	00
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). $40.00 per property +				$	
TOTAL FEES ENCLOSED =				$ 2358	00
				Amount to be refunded:	$
				charged:	$ 2358.00

a. ☒ A check in the amount of $ 2358.00 to cover the above fees is enclosed.

b. ☐ Please charge my Deposit Account No. _____ in the amount of $ _____ to cover the above fees. A duplicate copy of this sheet is enclosed.

c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 02-4550. A duplicate copy of this sheet is enclosed.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137 (a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

William D. Noonan, M.D.
Klarquist Sparkman Campbell Leigh & Whinston, LLP
1 World Trade Center, Suite 1600
121 SW Salmon Street
Portland, Oregon 97204

William D Noonan
SIGNATURE

William D. Noonan, M.D.
NAME

30,878
REGISTRATION NUMBER

WDN:can 8/21/98 4239-50420

PATENT

Express Mail No. EL121478350US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of

Meltzer et al.

Date of Deposit: August 21, 1998

For: AIB1, A NOVEL STEROID RECEPTOR COACTIVATOR

Date: August 21, 1998

CERTIFICATE OF MAILING

I HEREBY CERTIFY THAT THIS PAPER AND THE DOCUMENTS REFERRED TO AS BEING ATTACHED OR ENCLOSED HEREWITH ARE BEING DEPOSITED WITH THE UNITED STATES POSTAL SERVICE ON August 21, 1998 AS EXPRESS MAIL NO. EM295379309US IN AN ENVELOPE ADDRESSED TO: BOX PCT, ASSISTANT COMMISSIONER FOR PATENTS, WASHINGTON, DC 20231.

William D Noonan

William D. Noonan, M.D.
Attorney for Applicant

PRELIMINARY AMENDMENT

ASSISTANT COMMISSIONER FOR PATENTS
WASHINGTON, DC 20231

Prior to calculating the fee for or examining the accompanying patent application, please enter the following amendments:

In the claims:

9. A substantially pure DNA comprising (a) the [sequence of SEQ. I.D. NO. 1] DNA of claim 7 or (b) a degenerate variant thereof.

14. A method of identifying a candidate compound which inhibits estrogen receptor (ER)-dependent transcription comprising contacting the compound with [an] the AIB1 polypeptide of claim 12 and determining whether the compound binds to the polypeptide, wherein binding of the compound to the polypeptide indicates that the compound inhibits ER-dependent transcription.

28. [A] The method of claim 22, further comprising a method of detecting breast cancer in a tissue sample, comprising determining the number of cellular copies of an AIB1 gene in the tissue sample, wherein an increase in the number of copies compared to the number of copies in a normal control tissue indicates the presence of a breast carcinoma.

36. [A] The method of claim 32, further comprising a method of inhibiting ER-dependent transcription in a breast cell of an mammal, comprising administering an effective amount of an AIB1 polypeptide to the mammal.

40. [A] The method of claim 32, further comprising a method of inhibiting ER-dependent transcription in a cancer cell of a mammal, comprising administering an effective amount of a peptide mimetic of an AIB1 polypeptide to the mammal.

46. [A transgenic mouse wherein] The transgenic animal of claim 45, wherein at least one copy of the pCIP gene has been functionally deleted.

53. [A method of reducing proliferation of a cancer cell comprising administering to the mammal a] The method of claim 32, wherein the compound which inhibits interaction of AIB1 [with] comprises a molecule selected from the group consisting of steroid receptors and nuclear co-factors.

Respectfully submitted,

KLARQUIST SPARKMAN CAMPBELL
LEIGH & WHINSTON, LLP

By William D. Noonan
William D. Noonan
Registration No. 30,878

One World Trade Center, Suite 1600
121 S.W. Salmon Street
Portland, Oregon 97204
Telephone: (503) 226-7391

cc: Joseph K. Hemby, Jr. (WO # 101-98)
Paul Meltzer, M.D., Ph.D.
Ronald King, Ph.D.

S:\WPDOCS\WDN\NIH\50420\Preliminary.amd

AIB1, A NOVEL STEROID RECEPTOR CO-ACTIVATOR

BACKGROUND OF THE INVENTION

Breast cancer arises from estrogen-responsive breast epithelial cells. Estrogen activity is thought to promote the development of breast cancer, and many breast cancers are initially dependent on estrogen at the time of diagnosis. Anti-estrogen compositions have therefore been used to treat breast cancer.

A frequent mechanism of increased gene expression in human cancers is amplification, i.e., the copy number of a DNA sequence is increased, in a cancer cell compared to a non-cancerous cell. In breast cancer, commonly amplified regions are derived from 17q21, 8q24, and 11q13 which encode erbB-2, c-myc, and cyclic D1 respectively (Devilee et al., 1994, Crit. Rev. Oncog. 5:247-270). Recently, molecular cytogenetic studies have revealed the occurrence in breast cancers of additional regions of increased DNA copy number (Isola et al., Am. J. Pathol. 147:905-911, 1995; Kallioniemi et al., Proc. Natl. Acad. Sci. USA 91:2156-2160, 1994; Muleris et al., Genes Chromo. Cancer 10:160-170, 1994; Tanner et al., Cancer Research 54:4257-4260, 1994; Guan et al., Nat. Genet. 8:155-161, 1994).

Breast cancer is the second leading cause of cancer deaths in American women, and it is estimated that an American woman has at least a 10% cumulative lifetime risk of developing this disease. Early diagnosis is an important factor in breast cancer prognosis and affects not only survival rate, but the range of therapeutic options available to the patient. For instance, if diagnosed early, a "lumpectomy" may be performed, whereas later diagnosis tends to be associated with more invasive and traumatic surgical treatments such as radical mastectomy. The treatment of other cancers likewise is benefitted by early diagnosis, for instance the prognosis in the treatment of lung cancer, colorectal cancer and prostate cancers is greatly improved by early diagnosis. There is a need for a simple and reliable method of diagnosis of cancers in general and of breast cancer in particular. There is a need for a method of screening for compounds that inhibit the interaction between an estrogen receptor ER and an ER-dependent nuclear receptor co-activator molecule in order to identify molecules useful in research diagnosis and treatment of cancer. There is also a need for a method for identifying tamoxifen-sensitive cancer patients in order to better manage treatment. A solution to these needs would improve cancer treatment and research and would save lives.

SUMMARY OF THE INVENTION

The inventors have discovered that the AIB1 protein (Amplified In Breast Cancer-1) is a member of the Steroid Receptor Coactivator - 1 (SRC-1) family of nuclear receptor co-activators that interacts with estrogen receptors (ER) to enhance ER-dependent transcription. The inventors have further discovered that the AIB1 gene is amplified and over-expressed in certain cancers including breast cancer, and that detection of amplified AIB1 genes can therefore be used to detect

cancerous cells. Importantly, the inventors have also found that AIB1 amplification is not confined to breast cancer but is also found in cancers of the lung, ovary, head and neck, colon, testicles, bladder, prostate, endometrium, kidney, stomach and also in pheochromocytoma, melanoma, ductal carcinoma and carcinoid tumor. Such a finding means that AIB1 may be useful in the detection and treatment of all of the aforementioned cancers which include some of the most prevalent and deadly diseases in the western world.

The inventors have also discovered that AIB1 interacts with the proteins p300 and CBP, which are nuclear cofactors that interact with other nuclear factors to promote transcription (Chacravarti et al., *Nature* (383) 99-103 1996; Lundblad et al., *Nature* (374) 85-88 1995). The inventors have, furthermore, determined that in cells with stable over-expression of AIB1, there is a dramatic increase in steroid receptor activation (almost a 100-fold increase) leading to a corresponding increase in transcriptional activation. The inventors have also used monoclonal anti-AIB1 antibodies to demonstrate that AIB1 gene amplification is directly correlated with increased AIB1 expression, and that these amplified copies of the gene are expressed in physiological conditions. The inventors have found that AIB1 is the human ortholog of the mouse ER-dependent transcriptional activator p/CIP, with the proteins having an overall amino acid identity of 81.6%. These finding support the physiological role for AIB1 in cancer cells as a cofactor involved in transcriptional regulation.

The invention features a substantially pure DNA which includes a sequence encoding an AIB1 polypeptide, e.g., a human AIB1 polypeptide, or a fragment thereof. The DNA may have the sequence of all or part of the naturally-occurring AIB1-encoding DNA or a degenerate variant thereof. AIB1-encoding DNA may be operably linked to regulatory sequences for expression of the polypeptide. A cell containing AIB1 encoding DNA is also within the invention.

The invention also includes a substantially pure DNA containing a polynucleotides which hybridizes at high stringency to a AIB1-encoding DNA or the complement thereof. A substantially pure DNA containing a nucleotide sequence having at least 50% sequence identity to the full length AIB1 cDNA, e.g., a nucleotide sequence encoding a polypeptide having the biological activity of a AIB1 polypeptide, is also included.

The invention also features a substantially pure human AIB1 polypeptide and variants thereof, e.g., polypeptides with conservative amino acid substitutions or polypeptides with conservative or non-conservative amino acid substitutions which retain the biological activity of naturally-occurring AIB1.

Diagnostic methods, e.g., to identify cells which harbor an abnormal copy number of the AIB1 DNA, are also encompassed by the invention. An abnormal copy number, e.g., greater than the normal diploid copy number, of AIB1 DNA is indicative of an aberrantly proliferating cell, e.g., a steroid hormone-responsive cancer cell.

The invention also includes antibodies, e.g., a monoclonal antibody or polyclonal antisera, which bind specifically to AIB1 and can be used to detect the level of expression of AIB1 in a cell

or tissue sample. An increase in the level of expression of AIB1 in a patient-derived tissue sample compared to the level in normal control tissue indicates the presence of a cell proliferative disorder such as cancer.

Screening methods to identify compounds which inhibit an interaction of AIB1 with a steroid hormone receptor, thus disrupting a signal transduction pathway which leads to aberrant cell proliferation, is also within the invention. Proliferation of a cancer cell can therefore be reduced by administering to an individual, e.g., a patient diagnosed with a steroid-responsive cancer, a compound which inhibits expression of AIB1.

The invention also includes a knockout mutant, for example a mouse (or other mammal) from which at least one AIB1 gene has been selectively deleted from its genome. Such a mouse is useful in research, for instance, the phenotype gives insight into the physiological role of the deleted gene. For instance the mutant may be defective in specific biochemical pathways; such a knockout mutant may be used in complementation experiments to determine the role of other genes and proteins to determine if any such genes or proteins complement for the deleted gene. Homozygous and heterozygous mutants are included in this aspect of the invention.

The present invention also includes a mutant organism, for example a mammal such as a mouse which contains more than the normal number of AIB1 genes in its genome. Such a mouse may contain additional copies of the AIB1 gene integrated into its chromosomes, for instance in the form of a pro-virus, or may carry additional copies on extra-chromosomal elements such as plasmids. Such a mutant mouse is useful for research purposes, to elucidate the physiological or pathological role of AIB1. For instance, the role of AIB1 expression as cause or effect in cancers may be investigated by including or transplanting tumors into such mutants, and comparing such mutants with normal mice having the same cancer.

The present invention also includes a mutant organism, for example a mammal, e.g. a mouse, that contains, either integrated into a chromosome or on a plasmid, at least one copy of the AIB1 gene driven by a non-native promoter. Such a promoter may be constitutive or may be inducible. For instance, the AIB1 gene may be operatively linked to a mouse mammary tumor virus (MMTV) promoter or other promoter from a mammalian virus allowing manipulation of AIB1 expression. Such a mutant would be useful for research purposes to determine the physiological or pathological role of AIB1. For instance, over or under expression could be affected and physiological effects observed.

The invention also includes methods for treatment of cancers that involve functions of or alterations in the signaling pathways that use p300 and/or CBP as signal transducing molecules. The treatments of the invention involve targeting of the AIB1 protein or AIB1 gene to enhance or reduce interaction with p300 and/or CBP proteins. For instance, the AIB1 gene sequence as disclosed herein may be used to construct an anti-sense nucleotide. An anti-sense RNA may be constructed that is anti-parallel and complementary to the AIB1 transcript (or part thereof) and which will therefore form an RNA-RNA duplex with the AIB1 transcript, preventing transcription

and expression of AIB1. Alternatively, treatments may comprise contacting an AIB1 protein with a molecule that specifically binds to the AIB1 molecule *in vivo*, thereby interfering with AIB1 binding with other factors such as p300 or CBP. Such processes are designed to inhibit signal transduction pathways involving AIB1, p300, CBP and other factors and therefore inhibit cancer cell proliferation that is effected via these pathways. As explained in more detail below, AIB1 overexpression results in increased ER-dependent transcriptional activity which confers a growth advantage upon AIB1 amplification-bearing clones during the development and progression of estrogen-dependent cancers.

Compounds which inhibit or disrupt the interaction of an AIB1 gene product with a steroid hormone receptor, e.g., ER, are useful as anti-neoplastic agents for the treatment of patients suffering from steroid hormone-responsive cancers such as breast cancer, ovarian cancer, prostate cancer, and colon cancer.

AIB1 polypeptides or peptide mimetics of such polypeptides, e.g., those containing domains which interact with steroid hormone receptors, can be administered to patients to block the interaction of endogenous intracellular AIB1 and a steroid hormone receptor, e.g., ER in an aberrantly proliferating cell. It is likely that AIB1 interacts with a wide range of human transcriptional factors and that regulation of such interactions will have important therapeutic applications.

Other features and advantages of the invention will be apparent from the following description of the preferred embodiments thereof, and from the claims.

SEQUENCE LISTING

The nucleic acid and amino acid sequences listed in the accompanying Sequence Listing are shown using standard letter abbreviations for nucleotide bases and three-letter code for amino acids. Only one strand of each nucleic acid sequence is shown, but the complementary strand is understood to be included by any reference to the displayed strand.

SEQ. I.D. No. 1 shows the nucleic acid sequence of the human AIB1 cDNA and the corresponding amino acid sequence.

SEQ. I.D. No. 2 shows the amino acid sequence of the Per/Arnt/Sim (PAS) domain of AIB1.

SEQ. I.D. No. 3 shows the amino acid sequence of the basic helix-loop-helix domain (bHLH) of AIB1.

SEQ. I.D. No. 4 shows the amino acid sequence of the human AIB1 protein.

SEQ. I.D. No. 5 shows the nucleic acid sequence of primer N8F1.

SEQ. I.D. No. 6 shows the nucleic acid sequence of the forward primer designed from the 5' sequence of pCMVSPORT-B11, PM-U2.

SEQ. I.D. No. 7 shows the nucleic acid sequence of the reverse primer designed from the 5' sequence of pCMVSPORT-B11, PM-U2.

SEQ. I.D. No. 8 shows the amino acid sequence of the ER-interacting domain of AIB1.

SEQ. I.D. No. 9 shows the nucleic acid sequence of pCIP, the mouse ortholog of AIB1 and the amino acid sequence for this gene.

SEQ. I.D. No. 10 shows the nucleic acid sequence of the forward primer AIB1/mESTF1 used to screen mouse BAC.

SEQ. I.D. No. 11 shows the nucleic acid sequence of the reverse primer AIB1/mESTR1 used to screen mouse BAC.

SEQ. I.D. No. 12 shows the amino acid sequence of pCIP, the mouse ortholog of AIB1.

FIGURES

Fig. 1A is a diagram of an amino acid sequence of full length AIB1 in which residues highlighted in black are identical in AIB1, TIF2 and SRC1. Residues identical with TIF2 (GenBank accession number X97674) or SRC-1 (GenBank accession number U59302) are highlighted in grey or boxed, respectively.

Fig. 1B is a diagram showing the structural features of AIB1. The following domains are indicated: bHLH domain, PAS domains (with the highly conserved PAS A and B regions shown in dark gray), S/T (serine/threonine)-rich regions, and a group of charged residues (+/-). A glutamine-rich region and polyglutamine tract are also indicated. The numbers beneath the diagram indicate the location (approximate residue number) of the domain with respect to the amino acid sequence shown in Fig. 1A. The alignment was generated using DNASTAR software.

Fig. 2 is a photograph of a Northern blot analysis showing increased expression of AIB1 in the cell lines BT-474, ZR-75-1, MCF7, and BG-1.

Fig. 3 is a bar graph showing that the addition of full length AIB1 DNA to a cell resulted in an increase of estrogen-dependent transcription from an ER reporter plasmid. COS-1 cells were transiently transfected with 250 ng ER expression vector (pHEGO-hyg), 10 ng of luciferase reporter plasmid (pGL3.luc.3ERE or 10 ng pGL3 lacking ERE) and increasing amounts of pcDNA3.1-AIB1 and incubated in the absence (open bars) or presence of 10 nM 17β-stradiol (E2, solid bars) or 100 nM 4-hydroxytamoxifen (hatched bars). Luciferase activity was expressed in relative luminescence units (RLU). The data are the mean of three determinations from one of four replicate experiments. Error bars indicate one standard deviation.

Fig. 4 is a schematic diagram comparing the DNA and protein structures of pCIP (the mouse ortholog of AIB1) and the human AIB1; exons are shown as black boxes.

Fig. 5 is a table showing the introns and exons of the mouse AIB1 gene (pCIP). The "Exon" column refers to the number of the exon; "cDNA bp 5'-exon" refers to the nucleotide position in the mouse cDNA sequence for the 5' exon. "3' intron splice cite" refers to the last few nucleotides of the 3' position of the intron. "Exon sequence" refers to the exon itself. "5' intron" refers to the adjacent intron reading from the exon into the splice donor elinucleotides (usually GT).

Fig. 6 is a table showing the introns and exons of the human AIB1 gene. The "Exon" column refers to the number of the exon; "cDNA bp 5'-exon" refers to the nucleotide position in the mouse cDNA sequence for the 5' exon. "3' intron splice cite" refers to the last few nucleotides of the 3' position of the intron. "Exon sequence" refers to the exon itself. "5' intron" refers to the adjacent intron reading from the exon into the splice donor nucleotides (usually GT).

DETAILED DESCRIPTION

The invention is based on the discovery of a novel gene, amplified in breast cancer-1 (AIB1), which is overexpressed in breast cancer. AIB1 has the structural features of a co-activator of the steroid hormone receptor family. The steroid hormone estrogen and other related steroid hormones act on cells through specific steroid receptors.

Members of the steroid receptor coactivator (SRC) family of transcriptional co-activators interact with nuclear hormone receptors to enhance ligand-dependent transcription. AIB1 is a novel member of the SRC family which was found to be overexpressed in breast cancers. The AIB1 gene is located at human chromosome 20q. High-level AIB1 amplification and overexpression were observed in several estrogen receptor (ER) positive breast and ovarian cancer cell lines, as well as in uncultured breast cancer specimens. AIB1 amplification is not confined to breast cancer but is also found in cancers of the lung, ovary, head and neck, colon, testicles, bladder, prostate, endometrium, kidney, stomach and also in pheochromocytoma, melanoma, ductal carcinoma and carcinoid tumor.

Transfection of AIB1 into cells resulted in marked enhancement of estrogen-dependent transcription. These observations indicated that AIB1 functions as a co-activator of steroid hormone receptors such as ER (including estrogen receptor α (ERα) and estrogen receptor β (ERβ)), androgen receptor (e.g., expressed in prostate cells), retinoid receptor (e.g., isoforms α, γ, and retinoid X receptor (RXR)), progesterone receptor (e.g., expressed in breast cells), mineralocorticoid receptor (implicated in salt metabolism disorders), vitamin D receptor (implicated in calcium metabolism disorders), thyroid hormone receptor (e.g, thyroid hormone receptor α), or glucocorticoid receptor (e.g., expressed in spleen and thymus cells). The altered expression of AIB1 contributes to the initiation and progression of steroid hormone-responsive cancers by increasing the transcriptional activity of the steroid receptor.

A substantially pure DNA which includes an AIB1-encoding polynucleotides (or the complement thereof) is claimed. By "substantially pure DNA" is meant DNA that is free of the genes which, in the naturally-occurring genome of the organism from which the DNA of the invention is derived, flank the AIB1 gene. The term therefore includes, for example, a recombinant DNA which is incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote at a site other than its natural site; or which exists as a separate molecule (e.g., a cDNA or a genomic or cDNA fragment produced by PCR or restriction endonuclease digestion) independent of other sequences. It also includes a

recombinant DNA which is part of a hybrid gene encoding an additional polypeptide sequence. Preferably, the polypeptide includes a Per/Arnt/Sim (PAS) domain (LLQALDGFLFVVNRDGNIVFVSENVTQYLQYKQEDLVNTSVYNILHEEDRKDFLKNLPKST VNGVSWTNETQRQKSHTFNCRMLMKTPHDILEDINASPEMRQRYETMQCFALSQPRAMME EGEDLQSCMICVARRITTGERTFPSNPESFITRHDLSGKVVNIDTNSLRSSMRPGFEDIIRRCIQ ; SEQ. I.D. NO. 2) and/or a basic helix-loop-helix (bHLH) domain (RKRKLPCDTPGQGLTCSGEKRRREQESKYIEELAELISANLSDIDNFNVKPD KCAILKETVRQIRQIKEQGKT; SEQ. I.D. NO. 3); more preferably, the AIB1 polypeptide includes the amino acid sequence of the entire naturally-occurring AIB1 protein (Fig. 1; SEQ. I.D. NO. 4). Preferably, the peptide includes an ER-interacting domain of AIB1 (e.g., a domain comprising approximately amino acids 300 to 1250:

CIQRFFSLNDGQSWSQKRHYQEAYLNGHAETPVYRFSLADGTIVTAQTKSKLF
RNPVTNDRHGFVSTHFLQREQNGYRPNPNPVGQGIRPPMAGCNSSVGGMSMS
PNQGLQMPSSRAYGLADPSTTGQMSGARYGGSSNIASLTPGPGMQSPSSYQNNNYGLNMSS
PPHGSPGLAPNQQNIMISPRNRGSPKIASHQFSPVAGVHSPMASSGNTGNHSFSSSSLSALQAI
SEGVGTSLLSTLSSPGPKLDNSPNMNITQPSKVSNQDSKSPLGFYCDQNPVESSMCQSNSRDH
LSDKESKESSVEGAENQRGPLESKGHKKLLQLLTCSSDDRGHSSLTNSPLDSSCKESSVSVTS
PSGVSSSTSGGVSSTSNMHGSLLQEKHRILHKLLQNGNSPAEVAKITAEATGKDTSSITSCGD
GNVVKQEQLSPKKKENNALLRYLLDRDDPSDALSKELQPQVEGVDNKMSQCTSSTIPSSSQE
KDPKIKTETSEEGSGDLDNLDAILGDLTSSDFYNNSISSNGSHLGTKQQVFQGTNSLGLKSSQ
SVQSIRPPYNRAVSLDSPVSVGSSPPVKNISAFPMLPKQPMLGGNPRMMDSQENYGSSMGGP
NRNVTVTQTPSSGDWGLPNSKAGRMEPMNSNSMGRPGGDYNTSLPRPALGGSIPTLPLRSN
SIPGARPVLQQQQQMLQMRPGEIPMGMGANPYGQAAASNQLGSWPDGMLSMEQVSHGTQ
NRPLLRNSLDDLVGPPSNLEGQSDERALLDQLHTLLSNTDATGLEEIDRALGIPELVNQGQA
LEPKQDAFQGQEAAVMMDQKAGLYGQTYPAQGPPMQGGFHLQGQSPSFNSMMNQMNQQ
GNFPLQGMHPRANIMRPRTNTPKQLRMQLQQRLQGQQFLNQSRQALELKMENPTAGGAA
VMRPMMQPQQGFLNAQMVAQRSRELLSHHFRQQRVAMMMQQQQQQQ (SEQ. I.D. NO. 8). A cell containing substantially purified AIB1-encoding DNA is also within the invention.

The invention also includes a substantially pure DNA which contains a polynucleotide which hybridizes at high stringency to an AIB1 cDNA having the sequence of SEQ. I.D. NO. 1, or the complement thereof and a substantially pure DNA which contains a nucleotide sequence having at least 50% (for example at least 75%, 90%,95%, or 98-100%) sequence identity to SEQ. I.D. NO. 1, provided the nucleotide sequence encodes a polypeptide having the biological activity of a AIB1 polypeptide. By "biological activity" is meant steroid receptor co-activator activity. For example, allelic variations of the naturally-occurring AIB1-encoding sequence (SEQ. I.D. NO. 1) are encompassed by the invention. Sequence identity can be determined by comparing the nucleotide sequences of two nucleic acids using the BLAST sequence analysis software, for instance, the

NCBI gapped BLAST 2.0 program set to default parameters. This software is available from The National Center for Biotechnology Information (www.ncbi.nlm,nih.gov/BLAST).

Hybridization is carried out using standard techniques such as those described in Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, (1989). "High stringency" refers to DNA hybridization and wash conditions characterized by high temperature and low salt concentration, e.g., wash conditions of 65° C at a salt concentration of approximately 0.1 X SSC. "Low" to "moderate" stringency refers to DNA hybridization and wash conditions characterized by low temperature and high salt concentration, e.g. wash conditions of less than 60° C at a salt concentration of at least 1.0 X SSC. For example, high stringency conditions may include hybridization at about 42°C, and about 50% formamide; a first wash at about 65°C, about 2X SSC, and 1% SDS; followed by a second wash at about 65°C and about 0.1% x SSC. Lower stringency conditions suitable for detecting DNA sequences having about 50% sequence identity to an AIB1 gene are detected by, for example, hybridization at about 42°C in the absence of formamide; a first wash at about 42°C, about 6X SSC, and about 1% SDS; and a second wash at about 50°C, about 6X SSC, and about 1% SDS.

A substantially pure DNA including (a) the sequence of SEQ ID NO. 1 or (b) a degenerate variant thereof is also within the invention. The AIB1-encoding DNA is preferably operably linked to regulatory sequences (including, e.g., a promoter) for expression of the polypeptide.

By "operably linked" is meant that a coding sequence and a regulatory sequence(s) are connected in such a way as to permit gene expression when the appropriate molecules (e.g., transcriptional activator proteins) are bound to the regulatory sequence(s).

The invention also includes a substantially pure human AIB1 polypeptide or fragment thereof. The AIB1 fragment may include an ER-interaction domain such as one having the amino acid sequence of SEQ. I.D. NO. 8. Alternatively, the fragment may contain the amino acid sequence of SEQ. I.D. NOS. 2, 3, or 4.

Screening methods to identify candidate compounds which inhibit estrogen-dependent transcription, AIB1 expression, or an AIB1/ER interaction (and as a result, proliferation of steroid hormone-responsive cancer cells) are within the scope of the invention. For example, a method of identifying a candidate compound which inhibits ER-dependent transcription is carried out by contacting the compound with an AIB1 polypeptide and determining whether the compound binds to the polypeptide. Binding of the compound to the polypeptide indicates that the compound inhibits ER-dependent transcription, and in turn, proliferation of steroid hormone-responsive cancer cells. Preferably, the AIB1 polypeptide contains a PAS domain or a bHLH domain. Alternatively, the method is carried out by contacting the compound with an AIB1 polypeptide and an ER polypeptide and determining the ability of the compound to interfere with the binding of the ER polypeptide with the AIB1 polypeptide. A compound which interferes with an AIB1/ER interaction inhibits ER-dependent transcription.

A method of screening a candidate compound which inhibits an interaction of an AIB1 polypeptide with an ER polypeptide in a cell includes the steps of (a) providing a GAL4 binding site linked to a reporter gene; (b) providing a GAL4 binding domain linked to either (i) an AIB1 polypeptide or (ii) an ER polypeptide; (c) providing a GAL4 transactivation domain II linked to the ER polypeptide if the GAL4 binding domain is linked to the AIB1 polypeptide or linked to the AIB1 polypeptide if the GAL4 binding domain is linked to the ER polypeptide; (d) contacting the cell with the compound; and (e) monitoring expression of the reporter gene. A decrease in expression in the presence of the compound compared to that in the absence of the compound indicates that the compound inhibits an interaction of an AIB1 polypeptide with the ER polypeptide.

Diagnostic methods to identify an aberrantly proliferating cell, e.g., a steroid hormone-responsive cancer cell such as a breast cancer cell, ovarian cancer cell, or prostate cancer cell, are also included in the invention. For example, a method of detecting an aberrantly proliferating cell in a tissue sample is carried out by determining the level of AIB1 gene expression in the sample. An increase in the level of gene expression compared to that in a normal control tissue indicates the presence of an aberrantly proliferating cell. AIB1 gene expression is measured using an AIB1 gene-specific polynucleotides probe, e.g. in a Northern assay or polymerase chain reaction (PCR)-based assay, to detect AIB1 mRNA transcripts. AIB1 gene expression can also be measured using an antibody specific for an AIB1 gene product, e.g., by immunohistochemistry or Western blotting.

Aberrantly proliferating cells, e.g., cancer cells, in a tissue sample may be detected by determining the number of cellular copies of an AIB1 gene in the tissue. An increase in the number of gene copies in a cell of a patient-derived tissue, compared to that in normal control tissue indicates the presence of a cancer. A copy number greater than 2 (the normal diploid copy number) is indicative of an aberrantly proliferative cell. Preferably, the copy number is greater than 5 copies per diploid genome, more preferably 10 copies, more preferably greater than 20, and most preferably greater than 25 copies. An increase in copy number compared to the normal diploid copy number indicates that the tissue sample contains aberrantly proliferating steroid hormone-responsive cancer cells. AIB1 copy number is measured by fluorescent *in situ* hybridization (FISH), Southern hybridization techniques, and other methods well known in the art (Kallioniemi et al., *PNAS* 91: 2156-2160 (1994); Guan et al., *Nature Genetics* 8: 155-161 (1994); Tanner et al., *Clin. Cancer Res.* 1: 1455-1461 (1995); Guan et al., *Cancer Res.* 56: 3446-3450 (August 1996); Anzick et al., *Science* 277: 965-968 (August 1997)).

Aberrantly proliferating cells can also be identified by genetic polymorphisms in the polyglutamine tract of AIB1, e.g., variations in the size of this domain which alter AIB1 co-activator activity.

The invention also includes methods of treating a mammal, e.g., a human patient. For example, a method of reducing proliferation of a steroid hormone-responsive cancer cell, e.g., an estrogen-responsive breast cancer cell, in a mammal is carried out by administering to the mammal a compound which inhibits expression of AIB1. The compound reduces transcription of AIB1-

encoding DNA in the cell. Alternatively, the compound reduces translation of an AIB1 mRNA into an AIB1 gene product in the cell. For example, translation of AIB1 mRNA into an AIB1 gene product is inhibited by contacting the mRNA with antisense polynucleotides complementary to the AIB1 mRNA.

A method of inhibiting ER-dependent transcription in a breast cell of a mammal is carried out by administering an effective amount of an AIB1 polypeptide or a peptide mimetic thereof to the mammal. Preferably, the polypeptide inhibits an AIB1/ER interaction; more preferably, the polypeptide contains an ER-interacting domain; a PAS domain or a bHLH domain of AIB1. By binding to ER, such a polypeptide inhibits binding of AIB1 to ER, thereby inhibiting ER-dependent transcription.

The invention also includes antibodies, e.g., a monoclonal antibody or polyclonal antisera, which bind specifically to AIB1. The term "antibody" as used in this invention includes whole antibodies as well as fragments thereof, such as Fab, Fab', $F(ab')_2$, and Fv which bind to an AIB1 epitope. These antibody fragments are defined as follows: (1) Fab, the fragment which contains a monovalent antigen-binding fragment of an antibody molecule produced by digestion of whole antibody with the enzyme papain to yield an intact light chain and a portion of one heavy chain; (2) Fab', the fragment of an antibody molecule obtained by treating whole antibody with pepsin, followed by reduction, to yield an intact light chain and a portion of the heavy chain; two Fab' fragments are obtained per antibody molecule; (3) $(Fab')_2$, the fragment of the antibody obtained by treating whole antibody with the enzyme pepsin without subsequent reduction; $F(ab')_2$, a dimer of two Fab' fragments held together by two disulfide bonds; (4) Fv, a genetically engineered fragment containing the variable region of the light chain and the variable region of the heavy chain expressed as two chains; and (5) single chain antibody ("SCA"), a genetically engineered molecule containing the variable region of the light chain, the variable region of the heavy chain, linked by a suitable polypeptide linker as a genetically fused single chain molecule. Methods of making these fragments are routine.

Also within the invention is a method of identifying a tamoxifen-sensitive patient (one who is likely to respond to tamoxifen treatment by a reduction in rate of tumor growth) wherein the method includes the steps of (a) contacting a patient-derived tissue sample with tamoxifen; and (b) determining the level of AIB1 gene expression or amplification in the sample. An increase in the level of expression or gene copy number compared to the level or cellular copy number in normal control tissue indicates that the patient is tamoxifen-sensitive.

AIB1 gene expression is measured using an AIB1 gene-specific polynucleotide probe, e.g., in a Northern blot or PCR-based assay to detect AIB1 mRNA transcripts or in a Southern blot or FISH assay to detect amplification of the gene (which correlates directly with AIB1 gene expression). Alternatively, AIB1 gene expression is measured by detecting an AIB1 gene product, e.g., using an AIB1-specific antibody.

Transgenic mammals, e.g., mice, which overexpress an AIB1 gene product, e.g., by virtue of harboring multiple copies of AIB1-encoding DNA, are also within the invention.

"Transgenic" as used herein means a mammal which bears a transgene, a DNA sequence which is inserted by artifice into an embryo, and which then becomes part of the genome of the mammal that develops from that embryo. Any non-human mammal which may be produced by transgenic technology is included in the invention; preferred mammals include, mice, rats, cows, pigs, sheep, goats, rabbits, guinea pigs, hamsters, and horses.

By "transgene" is meant DNA which is partly or entirely heterologous (i.e., foreign) to the transgenic mammal, or DNA homologous to an endogenous gene of the transgenic mammal, but which is inserted into the mammal's genome at a location which differs from that of the natural gene.

Also within the invention is a knockout mutant, for instance a knockout mouse wherein the mouse has had at least one copy of the AIB1 gene (also called the pCIP gene in mice) deleted from its genome. Such a knockout mutant would be useful in research, for instance the phenotype gives insight into the physiological role of AIB1. Complementation experiments using such a knockout mutant can be used to identify other genes and proteins that make up for the lack of AIB1 in the mutant to restore wild-type phenotype.

Also within the invention is a mutant, such as a mouse, which contains more than the normal number of copies of the AIB1 (pCIP) gene, either integrated into a chromosome, for instance as a pro-virus, or in an extra-chromosomal element, such as on a plasmid.

Also within the invention is a mutant, for example, a mouse, which contains the AIB1 (pCIP) gene driven by a non-native promoter, such as a constitutive or an inducible promoter, such as the mouse mammary tumor virus (MMTV) promoter.

The invention also includes methods of treatment for cancers the growth of which involves alternations of signaling pathways involving p300 and/or CBP. For example, AIB1 (pCIP) may be contacted with a molecule that binds to AIB1 and inhibits AIB1's interaction with p300, thereby disrupting signaling of this pathway and reducing transcription of molecules whose transcription is positively regulated by this pathway; thereby reducing tumor growth.

Example 1: Cloning and Expression of AIB1

A. Cloning of AIB1

Chromosome microdissection and hybrid selection techniques were used to isolate probes and clone gene sequences which map to chromosome 20q, one of the recurrent sites of DNA amplification in breast cancer cells identified by molecular cytogenetics (Kallioniemi et al., *PNAS* 91: 2156-2160 (1994); Guan et al., *Nature Genetics* 8: 155-161 (1994); Tanner et al., *Clin. Cancer Res.* 1: 1455-1461 (1995); Guan et al., *Cancer Res.* 56: 3446-3450 (August 1996); Anzick et al., *Science* 277: 965-968 (August 1997)). AIB1 is a member of the SRC-1 family of nuclear receptor (NR) co-activators. AIB1 functions to enhance ER-dependent transcription. SRC-1 and the closely

related TIF2 are steroid receptor co-activators with an affinity for NRs. The mouse ortholog of human AIB1 is called pCIP. In this application pCIP and AIB1 will be used synonymously unless the contrary is clearly expressed.

To characterize AIB1, the full length cDNA was cloned and sequenced. An AIB1 specific primer N8F1 (5'-TCATCACTTCCGACAACAGAGG-3'; SEQ. I.D. NO. 5) was biotinylated and used to capture cDNA clones from a human lung cDNA library (Gibco, BRL) using the GENETRAPPER cDNA Positive Selection System (Gibco, BRL). The largest clone (5.8 kb), designated pCMVSPORT-B11, was selected for sequence analysis. To obtain full-length AIB1-encoding DNA, a random-primed library from BT-474 was constructed in bacteriophage λ-Zap (Stratagene) and hybridized with a 372 bp ^{32}P-labeled PCR product amplified from a human spleen cDNA library using primers designed form the 5' sequence of pCMVSPORT-B11, PM-U2 (5'-CCAGAAACGTCACTATCAAG-3', forward primer; SEQ. I.D. NO. 6) and B11-11RA (5'-TTACTGGAACCCCCATACC-3', reverse primer; SEQ. I.D. NO. 7). Plasmid rescue of 19 positive clones yielded a clone, pBluescript-R22, which overlapped pCMVSPORT-B11 and contained the 5' end of the coding region. To generate a full length AIB1 clone, the 4.85 kb HindIII/XhoI fragment of pCMVSPORT-B11 was subcloned into HindIII/XhoI sites of pBluescript-R22. The 4.84 kb NotI/NheI fragment of the full length clone containing the entire coding region was then subcloned into the NotI/XbaI sites of the expression vector, pcDNA3.1 (Invitrogen), generating pcDNA3.1-AIB1.

The cloned DNA sequence (SEQ. I.D. No. 1) revealed an open reading frame (beginning at the underlined "ATG") encoding a protein of 1420 amino acids with a predicted molecular weight of 155 kDa (Fig. 1A). Database searches with BLASTP identified a similarity of AIB1 with TIF2 (45% protein identity) and SRC-1 (33% protein identity). Like TIF2 and SRC-1, AIB1 contains a bHLH domain preceding a PAS domain, serine/threonine-rich regions, and a charged cluster (Fig. 1B). There is also a glutamine-rich region which, unlike SRC-1 and TIF2, contains a polyglutamine tract (Fig. 1B). The polyglutamine tract of AIB1 is subject to genetic polymorphism. Variations in the size of this domain alter AIB1 co-activator activity.

B. Expression of AIB1

Amplification and expression of AIB1 in several ER positive and negative breast and ovarian cancer cell lines was examined. Established breast cancer cell lines used in the experiments described below (see, e.g., Fig. 2) were obtained from the American Type Culture Collection (ATCC): BT-474, MCF-7, T-47D, MDA-MB-361, MDA-MB-468, BT-20, MDA-MB-436, and MDA-MB-453; the Arizona Cancer Center (ACC): UACC-812; or the National Cancer Institute (NCI): ZR75-1.

AIB1 gene copy number was determined by FISH. For FISH analysis, interphase nuclei were fixed in methanol:acetic acid (3:1) and dropped onto microscope slides. AIB1 amplification was detected in the breast cancer cell line ZR75-1, the ovarian cancer cell line BG-1, and two

uncultured breast cancer samples. Intra-chromosomal amplification of AIB1 was apparent in metaphase chromosomes of ZR75-1 and BG1. Numerous copies of AIB1 were resolved in the adjacent interphase nuclei. Extrachromosomal copies (e.g., in episomes or double minute chromosomes) of AIB1 have also been detected. The Spectrum-Orange (Vysis) labeled AIB1 P1 probe was hybridized with a biotinylated reference probe for 20q11 (RMC20P037) or a fluorescein labeled probe for 20p (RMC20C039).

High level amplification of AIB1 (greater than 20 fold), similar to that observed in BT-474 and MCF-7, was seen in two additional ER-positive cell lines, breast carcinoma ZR75-1, and ovarian carcinoma BG-1 (see Fig. 2). Interphase FISH studies demonstrated that amplification of chromosome 20q in breast cancer is complex, involving several distinct variably co-amplified chromosomal segments derived from 20q11, 20q12, and 20q13. Probes for the 20q11 and 20q13 regions of amplification did not detect amplification in ZR75-1 and BG-1, suggesting that amplification of AIB1 (which maps to 20q12) occurred independently in these cell lines.

To determine if AIB1 amplification also occurred in uncultured cells from patient biopsies, breast cancer specimens were screened for AIB1 amplification by interphase FISH. In two of 16 specimens analyzed, high AIB1 copy number (up to 25 copies/cell) was detected. Both tumor specimens tested came from post-menopausal patients and were ER/PR positive. One of the specimens was obtained from a metastatic tumor of a patient who subsequently responded favorably to tamoxifen treatment.

AIB1 expression was also examined in cells with and without AIB1 amplification and compared to expression of ER, SRC-1 and TIF2 by Northern blotting. In accordance with its amplification status, AIB1 was highly overexpressed in BT-474, MCF-7, ZR75-1, and BG-1 (Fig. 2). Three of the four cell lines exhibiting AIB1 overexpression also demonstrated prominent ER expression, while two others displayed lower but detectable ER expression (BT-474 and BT-20). Fig. 2 also shows that the expression of TIF2 and SRC-1 remained relatively constant in all cell lines tested. Taken together, these observations demonstrate that AIB1 amplification is associated with significant overexpression of AIB1 gene product. The correlation of elevated AIB1 expression with ER positivity in tumors indicates that AIB1 is a component of the estrogen signaling pathway, the amplification of which is selected during cancer development and progression.

To determine whether expression of AIB1 increases ER ligand-dependent transactivation, transient transfection assays were performed. The effect of increasing levels of AIB1 on transcription of an ER dependent reporter was measured. The results demonstrated that co-transfection of AIB1 led to a dose dependent increase in estrogen-dependent transcription (Fig. 3). This effect was not observed when the estrogen antagonist, 4-hydroxytamoxifen (4-OHT), was substituted for 17β-estradiol or when the estrogen response element (ERE) was removed from the reporter plasmid (Fig. 3). A modest increase in basal transcription levels was observed with higher concentrations of AIB1 even in the absence of an ERE suggesting that AIB1 may have an intrinsic

transactivation function. These results demonstrate that, like the closely related TIF2 and SRC-1, AIB1 functions as an ER co-activator.

Example 2: Characterization of AIB1

A. **Functional Domains of AIB1**

TIF-2, SRC-1, and AIB1 are characterized by highly conserved N-terminal bHLH and PAS domains. The PAS region functions as a protein dimerization interface in the mammalian aryl hydrocarbon receptor and the aryl hydrocarbon receptor nuclear transporter proteins, as well as the *Drosophila* transcription factors *sim* and *per*. The PAS region (SEQ. I.D. NO. 2) of AIB1 functions as a protein interaction domain, mediating binding between AIB1 and other proteins. However, steroid hormone activators lacking the PAS domain are capable of interacting with nuclear steroid hormone receptors. The highly conserved bHLH domain (SEQ. I.D. NO. 3) participates in protein interactions which mediate or modulate transmission of the hormone signal to the transcriptional apparatus. The ER-interacting domain (SEQ. I.D. NO. 8) mediates binding of AIB1 with a steroid hormone receptor protein.

AIB1 also interacts with the transcriptional integrators CREB binding protein (CBP) and p300. These transcriptional integrators interact directly with the basal transcriptional machinery. The CBP/p300 receptor association domain of AIB1 does not encompass the bHLH/PAS regions.

B. Purification of Gene Products

DNA containing a sequence that encodes part or all of the amino acid sequence of AIB1 can be subcloned into an expression vector, using a variety of methods known in the art. The recombinant protein can then be purified using standard methods. For example, a recombinant polypeptide can be expressed as a fusion protein in procaryotic cells such as *E. coli*. Using the maltose binding protein fusion and purification system (New England Biolabs), the cloned human cDNA sequence is inserted downstream and in frame of the gene encoding maltose binding protein (malE). The malE fusion protein is overexpressed in *E. coli* and can be readily purified in quantity. In the absence of convenient restriction sites in the human cDNA sequence, PCR can be used to introduce restriction sites compatible with the pMalE vector at the 5' and 3' end of the cDNA fragment to facilitate insertion of the cDNA fragment into the vector. Following expression of the fusion protein, it can be purified by affinity chromatography. For example, the fusion protein can be purified by virtue of the ability of the maltose binding protein portion of the fusion protein to bind to amylase immobilized on a column.

To facilitate protein purification, the pMalE plasmid contains a factor Xa cleavage site upstream of the site into which the cDNA is inserted into the vector. Thus, the fusion protein purified as described above can be cleaved with factor Xa to separate the maltose binding protein portion of the fusion protein from recombinant human cDNA gene product. The cleavage products can be subjected to further chromatography to purify recombinant polypeptide from the maltose binding protein. Alternatively, an antibody specific for the desired recombinant gene product can

be used to purify the fusion protein and/or the gene product cleaved from the fusion protein. Many comparable commercially available fusion protein expression systems can be utilized similarly.

AIB1 polypeptides can also be expressed in eucaryotic cells, e.g., yeast cells, either alone or as a fusion protein. For example, a fusion protein containing the GAL4 DNA-binding domain or activation domain fused to a functional domain of AIB1, e.g., the PAS domain, the bHLH domain, or the ER-interacting domain, can be expressed in yeast cells using standard methods such as the yeast two hybrid system described below. Alternatively, AIB1 polypeptides can be expressed in COS-1 cells using methods well known in the art, e.g., by transfecting a DNA encoding an AIB1 polypeptide into COS-1 cells using, e.g., the Lipofectamine transfection protocol described below, and culturing the cells under conditions suitable for protein expression.

Example 3: Detection of AIB1

A. Detection of Nucleotides Encoding AIB1

Determination of gene copy number in cells of a patient-derived sample is known in the art. For example, AIB1 amplification in cancer-derived cell lines as well as uncultured breast cancer cells was carried out using bicolor FISH analysis as follows. A genomic P1 clone containing AIB1 was labeled with Spectrum Orange-dUTP (Vysis) using the BioPrime DNA Labeling System (Gibco BRL). A 20q11 P1 clone was labeled with Biotin-16-dUTP (BMB) using nick translation. Fluorescent images were captured using a Zeiss axiophot microscope equipped with a CCD camera and IP Lab Spectrum software (Signal Analytics). Interphase FISH analysis of uncultured breast cancer samples was performed using known methods (Kallioniemi et al., *PNAS* 91: 2156-2160 (1994); Guan et al., *Nature Genetics* 8: 155-161 (1994); Tanner et al., *Clin. Cancer Res.* 1: 1455-1461 (1995); Guan et al., *Cancer Res.* 56: 3446-3450 (August 1996); Anzick et al., *Science* 277: 965-968 (August 1997)). Alternatively, standard Southern hybridization techniques can be employed to evaluate gene amplification. For example, Southern analysis is carried out using a non-repetitive fragment of genomic AIB1 DNA, e.g., derived from the 20q11 P1 clone described above or another AIB1 gene-containing genomic clone, as a probe.

The level of gene expression may be measured using methods known in the art, e.g., *in situ* hybridization, Northern blot analysis, or Western blot analysis using AIB1-specific monoclonal or polyclonal antibodies. AIB1 gene transcription was measured using Northern analysis. For example, the data shown in Fig. 2 was obtained as follows. The blot was hybridized sequentially with a probe (ER, AIB1, TIF2, SRC-1, or β-actin as indicated to the left of the photograph). AIB1 expression was compared to that of ER, TIF2, and SRC-1. cDNA clones were obtained from Research Genetics [TIF2 (clone 132364, GenBank accession no. R25318); SRC-1 (clone 418064, GenBank accession no. W90426)], the American Type Culture Collection (pHEGO-hyg, ATCC number 79995), or Clontech (β actin). The AIB1 probe was a 2.2kb NotI/SacI fragment of pCMVSPORT-B11. The β-actin probe was used as a control for loading error. To avoid cross-hybridization between these related genes and to match signal intensities, similar sized probes from

the 3'UTRs of AIB1, TIF2, and SRC-1 were utilized. Each of these probes detected a signal in normal mammary RNA on longer exposure. Electrophoresis, transfer and hybridization of 15 μg total RNA was performed by standard methods.

B. Detection of AIB1 Gene Products

AIB1 polypeptides to be used as antigens to raise AIB1-specific antibodies can be generated by methods known in the art, e.g., proteolytic cleavage, *de novo* synthesis, or expression of a recombinant polypeptide from the cloned AIB1 gene or a fragment thereof. AIB1-specific antibodies are then produced using standard methodologies for raising polyclonal antisera and making monoclonal antibody-producing hybridoma cell lines (see Coligan et al., eds., *Current Protocols in Immunology*, 1992, Greene Publishing Associates and Wiley-Interscience). To generate monoclonal antibodies, a mouse is immunized with an AIB1 polypeptide, antibody-secreting B cells isolated from the mouse, and the B cells immortalized with a non-secretory myeloma cell fusion partner. Hybridomas are then screened for production of an AIB1-specific antibody and cloned to obtain a homogenous cell population which produces a monoclonal antibody.

For administration to human patients, antibodies, e.g., AIB1 specific monoclonal antibodies, can be humanized by methods known in the art. Antibodies with a desired binding specificity can be commercially humanized (Scotgene, Scotland; Oxford Molecular, Palo Alto, CA).

Example 4: Detection of AIB1-related cell proliferative disorders

A. Diagnostic and Prognostic Methods

The invention includes a method of detecting an aberrantly proliferating cell, e.g., a steroid hormone-responsive cancer cell such as a breast cancer cell, an ovarian cancer cell, colon cancer cell, or prostate cancer cell, by detecting the number of AIB1 gene copies in the cell and/or the level of expression of the AIB1 gene product. AIB1 gene amplification or gene expression in a patient-derived tissue sample is measured as described above and compared to the level of amplification or gene expression in normal non-cancerous cells. An increase in the level of amplification or gene expression detected in the patient-derived biopsy sample compared to the normal control is diagnostic of a diseased state, i.e., the presence of a steroid hormone responsive cancer.

Because of the importance of estrogen exposure to mammary carcinogenesis and of anti-estrogen treatment in breast cancer therapy, such assays are also useful to determine the frequency of alterations of AIB1 expression in pre-malignant breast lesions (e.g. ductal carcinoma *in situ*) and during the progression from hormone dependent to hormone independent tumor growth.

The diagnostic methods of the invention are useful to determine the prognosis of a patient and estrogen responsive status of a steroid hormone-responsive cancer.

AIB1 expression can also be measured at the protein level by detecting an AIB1 gene products with an AIB1-specific monoclonal or polyclonal antibody preparation.

B. Diagnosis of Tamoxifen-Sensitivity

Overexpression of AIB1, e.g., as a result of AIB1 gene amplification, in steroid hormone-responsive cancers can predict whether the cancer is treatable with anti-endocrine compositions, e.g., tamoxifen. AIB1 amplification or overexpression in a patient-derived tissue sample compared to a normal (non-cancerous) tissue indicates tumor progression.

Absence of AIB1, e.g., loss of all or part of the AIB1 gene, but retention of ER-positivity in steroid hormone-responsive cancers predicts failure or poor responsiveness to anti-endocrine therapy, e.g., administration of anti-estrogen compositions such as tamoxifen. Since loss of AIB1 expression in a cancer cell may indicate a disruption of the ER signal transduction pathway, anti-estrogen therapy may be ineffective to treat such cancers. Patients identified in this manner (who would otherwise be treated with anti-estrogens) would be treated with alternative therapies.

Loss of estrogen receptor in recurrent breast caner is also associated with poor response to endocrine therapy. Up to 30% to 40% of metastases from hormone receptor-positive primary breast cancer do not respond to endocrine therapy. The frequency of hormone receptor status changes between primary and recurrent tumors and whether such a change might explain unresponsiveness to endocrine therapy was examined. Primary breast cancer samples and matched asynchronous recurrences were studied from 50 patients who had not received any adjuvant therapy. ER and progesterone receptor (PR) status was determined immunohistochemically from histologically representative formalin-fixed paraffin-embedded tumor samples. ER status was ascertained by mRNA in situ hybridization. Thirty-five (70%) of 50 primary tumors were positive for ER and 30 (60%) for PR. Hormone receptor status of the recurrent tumor differed from that of the primary tumor in 18 cases (36%). Discordant cases were due to the loss of ER (n=6), loss of PR (n=6), or loss of both receptors (n=6). Receptor-negative primary tumors were always accompanied by receptor-negative recurrences. Among 27 patients with ER-positive primary tumors, loss of ER was a significant predictor (P=.0085) of poor response to subsequent endocrine therapy. Only one of eight patients (12.5%) with lost ER expression responded to tamoxifen therapy, whereas the response rate was 74% (14 of 19) for patients whose recurrent tumors retained ER expression. Loss of ER expression in recurrent breast cancer predicts poor response to endocrine therapy in primarily ER-positive patients. Evaluation of ER expression and/or AIB1 expression (or gene copy number) is useful to determine the most effective approach to treatment of steroid-responsive cancers.

Example 5: Screening of candidate compounds

A. *In vitro* assays

The invention includes methods of screening to identify compounds which inhibit the interaction of AIB1 with ER, thereby decreasing estrogen dependent transcription which leads to aberrant cell proliferation. A transcription assay is carried out in the presence and absence of the candidate compound. A decrease in transcription in the presence of the compound compared to that

in its absence indicates that the compound blocks an AIB1/ER interaction and inhibits estrogen dependent transcription.

To determine the effect of AIB1 on estrogen-dependent transcription, an ER reporter plasmid can be used. The transcription assays described herein were conducted as follows. COS-1 cells were grown and maintained in phenol-red free DMEM medium supplemented with 10% charcoal-stripped fetal bovine serum. Cells were plated into 6-well culture dishes at 1.5 X 10^5 cells/well and allowed to grow overnight. Transfection of cells with the ER reporter plasmid was performed with Lipofectamine (Gibco, BRL) following the manufacturer's protocol. Three ng pRL-CMV were used as an internal control for transfection efficiency. Ligand or ethanol vehicle was added 234 hours post-transfection and cell lysates were harvested 48 hours post-transfection. Reporter activities were determined using the Dual-Luciferase Reporter Assay System (Promega) and the results expressed in relative luminescence units (RLU; luciferase/*Renilla* luciferase). pRL-CMV and pGL3-promoter were obtained from Promega. pHEGO-hyg was obtained from ATCC. The ER reporter pGL3.luc.3ERE contains three tandem copies of the ERE upstream from the SV40 promoter driving the luciferase gene. Standard mammalian expression vectors were utilized. Empty pcDNA3 vector was added to each of the pcDNA3.1-AIB1 dilutions to maintain constant amounts of plasmid DNA.

Compounds which inhibit the interaction of AIB1 with ER are also identified using a standard co-precipitation assay. AIB1/ER co-precipitation assays are carried out as follows. An AIB1 polypeptide and an ER polypeptide are incubated together to allow complex formation. One of the polypeptides is typically a fusion protein, e.g., GST-AIB1, and the other is tagged with a detectable label, e.g., ^{32}P-labeled ER). After incubation, the complex is precipitated, e.g., using glutathione-Sepharose beads. The beads are washed, filtered through a glass fiber filter, and collected. The amount of co-precipitated ^{32}P-label is measured. A reduction in the amount of co-precipitated label in the presence of a candidate compound compared to that in the absence of the candidate compound indicates that the compound inhibits an AIB1/ER interaction

Alternatively, a standard *in vitro* binding assay can be used. For example, one polypeptide, e.g., AIB1, can be bound to a solid support and contacted with the second polypeptide, e.g., ER. The amount of the second polypeptide which is retained on the solid support is then measured. A reduction in the amount of retained (second) polypeptide in the presence of a candidate compound compared to that in its absence indicates that the compound inhibits an AIB1/ER interaction. Techniques for column chromatography and coprecipitation of polypeptides are well known in the art.

An evaluation of AIB1/ER interaction and identification of compounds that blocks or reduces the interaction can also be carried out *in vivo* using a yeast two-hybrid expression system in which the activity of a transcriptional activator is reconstituted when the two proteins or polypeptides of interest closely interact or bind to one another.

The yeast *GAL4* protein consists of functionally distinguishable domains. One domain is responsible for DNA-binding and the other for transcriptional activation. In the two-hybrid expression system, plasmids encoding two hybrid proteins, a first fusion protein containing the *GAL4* DNA-binding domain fused to a first protein, e.g., AIB1, and the second fusion protein containing the *GAL4* activation domain fused to a second protein, e.g., ER, are introduced into yeast. If the two proteins are able to interact with one another, the ability to activate transcription from promoters containing Gal4-binding sites upstream from an activating sequence from GAL1 (UAS_G) is reconstituted leading to the expression of a reporter gene. A reduction in the expression of the reporter gene in the presence of a candidate compound compared to that in the absence of the compound indicates that the compound reduces an AIB1/ER interaction.

A method of identifying a DNA-binding protein which regulates AIB1 transcription can be carried out as follows:

A DNA containing a cis-acting regulatory element can be immobilized on polymeric beads, such as agarose or acrylamide. A mixture of proteins, such as a cell lysate, is allowed to come in contact with and bind to the DNA. Following removal of non-binding proteins, specifically-bound proteins, are eluted with a competing DNA sequence which may be identical to the immobilized sequence. Specific binding of a protein to the DNA regulatory element indicates that the protein may regulate AIB1 transcription. Functional activity of the identified trans-acting factor can be confirmed with an appropriate functional assay, such as one which measures the level of transcription of a reporter gene having the cis-acting regulatory gene 5' to the transcription start site of AIB1.

A method of identifying a compound which decreases the level of AIB1 transcription can be accomplished by contacting an immobilized AIB1-derived cis-acting regulatory element with a trans-acting regulatory factor in the presence and absence of candidate compound. A detectable change, i.e., a reduction, in specific binding of the trans-acting factor to its DNA target indicates that the candidate compound inhibits AIB1 transcription.

In addition to interacting with ER, AIB1 also interacts with the transcriptional integrators CBP and p300. CBP and p300 participate in the basal transcriptional apparatus in a cell. Thus, another approach to inhibit signal transduction through AIB1 is to prevent the formation of or disrupt an interaction of AIB1 with CBP and/or p300. Compounds which inhibit signal transduction (and therefore cell proliferation) can be identified by contacting AIB1 (or a fragment thereof which interacts with CBP or p300) with CBP or p300 (or a fragment thereof containing an AIB1-interacting domain, e.g., a C-terminal fragment) in the presence and absence of a candidate compound. For example, a C-terminal fragment of CBP involved in steroid receptor co-activator interaction contains 105 amino acids in the Q-rich region of CBP (Kamei et al., 1996, Cell 85:403-414; Yao et al., 1996, Proc. Natl. Acad. Sci. USA 93:10626-10631; Hanstein et al., 1996, Proc. Natl. Acad. Sci. USA 93:11540-11545). A decrease in AIB1 interaction with CBP or p300 in the presence of a candidate compound compared to that its absence indicates that the compound inhibits AIB1 interaction with these transcriptional integrators, and as a result, AIB1-mediated signal

transduction leading to DNA transcription and cell proliferation. Compounds which inhibit AIB1 interaction with transcriptional integrators can also be identified using a co-precipitation assay and the yeast two-hybrid expression system described above.

B. ***In vivo*** **assays**

Transgenic mice are made by standard methods, e.g., as described in Leder et al., U.S. Patent No. 4,736,866, herein incorporated by reference, or Hogan et al., 1986 Manipulating the Mouse Embryo. Cold Spring Harbor Laboratory" New York.

Briefly, a vector containing a promoter operably linked to AIB1-encoding cDNA is injected into murine zygotes, e.g., C57BL/6J X DBA/2F2 zygotes. Incorporation of the transgene into murine genomic DNA is monitored using methods well known in the art of molecular biology, e.g., dot blotting tail DNA with a probe complimentary to the 3' region of the gene contained in the AIB1 transgene construct. Mice thus confirmed to harbor the transgene can then be used as founders. Animal lines are created by crossing founders with C57BL/6J mice (The Jackson Laboratory, Bar Harbor, ME). AIB1 transgenic mice can be used to screen candidate compounds *in vivo* to identify compounds which inhibit aberrant cell proliferation, e.g., as measured by reduction tumor growth or metastasis. AIB1 transgenic mice are also useful to identify other genes involved in steroid hormone receptor-dependent cancers and to establish mouse cell lines which overexpress AIB1. AIB1-overexpressing cell lines are useful to screen for compounds that interfere with AIB1 function, e.g, by blocking the interaction of AIB1 with a ligand.

Example 6: AIB1 therapy

As discussed above, AIB1 is a novel member of the SRC-1 family of transcriptional co-activators. Amplification and overexpression of AIB1 in ER-positive breast and ovarian cancer cells and in breast cancer biopsies implicate this protein as a critical component of the estrogen response pathway. AIB1 overexpression results in increased ER-dependent transcriptional activity which confers a growth advantage of AIB1 amplification-bearing clones during the development and progression of estrogen-dependent cancers.

Compounds which inhibit or disrupt the interaction of an AIB1 gene product with a steroid hormone receptor, e.g., ER, are useful as anti-neoplastic agents for the treatment of patients suffering from steroid hormone-responsive cancers such as breast cancer, ovarian cancer, prostate cancer, and colon cancer. Likewise, compounds which disrupt interaction between AIB1 and p300 and/or CBP are also useful as anti-neoplastic agents.

AIB1 polypeptides or peptide mimetics of such polypeptides, e.g., those containing domains which interact with steroid hormone receptors, can be administered to patients to block the interaction of endogenous intracellular AIB1 and a steroid hormone receptor, e.g., ER in an aberrantly proliferating cell. A mimetic may be made by introducing conservative amino acid substitutions into the peptide. Certain amino acid substitutions are conservative since the old and

the new amino acid share a similar hydrophobicity or hydrophylicity or are similarly acidic, basic or neutrally charged (Stryer "Biochemistry" 1975, Ch.2, Freeman and Company, New York). Conservative substitutions replace one amino acid with another amino acid that is similar in size, hydrophobicity, etc. Examples of conservative substitutions are shown in the table below (Table 1).

TABLE 1

Original Residue	Conservative Substitutions
Ala	ser
Arg	lys
Asn	gln, his
Asp	glu
Cys	ser
Gln	asn
Glu	asp
Gly	pro
His	asn; gln
Ile	leu, val
Leu	ile; val
Lys	arg; gln; glu
Met	leu; ile
Phe	met; leu; tyr
Ser	thr
Thr	ser
Trp	tyr
Tyr	trp; phe
Val	ile; leu

Variations in the cDNA sequence that result in amino acid changes, whether conservative or not, should be minimized in order to preserve the functional and immunologic identity of the encoded protein.

Compositions administered therapeutically include polypeptide mimetics in which one or more peptide bonds have been replaced with an alternative type of covalent bond which is not susceptible to cleavage by peptidases. Where proteolytic degradation of the peptides following injection into the subject is a problem, replacement of a particularly sensitive peptide bond with a noncleavable peptide mimetic yields a more stable and thus more useful therapeutic polypeptide. Such mimetics, and methods of incorporating them into polypeptides, are well known in the art. Similarly, the replacement of an L-amino acid residue with a D-amino acid residue is a standard way of rendering the polypeptide less sensitive to proteolysis. Also useful are amino-terminal blocking groups such as t-butyloxycarbonyl, acetyl, theyl, succinyl, methoxysuccinyl, suberyl, adipyl, azelayl, dansyl, benzyloxycarbonyl, fluorenylmethoxycarbonyl, methoxyazelayl, methoxyadipyl, methoxysuberyl, and 2,4,-dinitrophenyl.

AIB1 polypeptides or related peptide mimetics may be administered to a patient, intravenously in a pharmaceutically acceptable carrier such as physiological saline. Standard methods for intracellular delivery of peptides can be used, e.g. packaged in liposomes. Such methods are well known to those of ordinary skill in the art. It is expected that an intravenous dosage of approximately 1 to 100 μmoles of the polypeptide of the invention would be administered per kg of body weight per day. The compositions of the invention are useful for parenteral administration, such as intravenous, subcutaneous, intramuscular, and intraperitoneal.

The therapeutic compositions of this invention may also be administered by the use of surgical implants which release the compounds of the invention. These devices could be readily implanted into the target tissue, e.g., a solid tumor mass, and could be mechanical or passive. Mechanical devices, such as pumps, are well known in the art, as are passive devices (e.g., consisting of a polymer matrix which contains therapeutic formulations; these polymers may slowly dissolve or degrade to release the compound, or may be porous and allow release via pores).

Antisense therapy in which a DNA sequence complementary to an AIB1 mRNA transcript is either produced in the cell or administered to the cell can be used to decrease AIB1 gene expression thereby inhibiting undesired cell proliferation, e.g., proliferation of steroid hormone-responsive cancer cells. An antisense polynucleotide, i.e., one which is complementary of the coding sequence of the AIB1 gene, is introduced into the cells in which the gene is overproduced. The antisense strand (either RNA or DNA) may be directly introduced into the cells in a form that is capable of binding to the transcripts. Alternatively, a vector containing a DNA sequence which, once within the target cells, is transcribed into the appropriate antisense mRNA, may be administered. An antisense nucleic acid which hybridizes to the coding strand of AIB1 DNA can decrease or inhibit production of an AIB1 gene product by associating with the normally single-stranded mRNA transcript, and thereby interfering with translation.

DNA is introduced into target cells of the patient with or without a vector or using standard vectors and/or gene delivery systems. Suitable gene delivery systems may include liposomes, receptor-mediated delivery systems, naked DNA, and viral vectors such as herpes viruses, retroviruses, and adenoviruses, among others. The DNA of the invention may be administered in a pharmaceutically acceptable carrier. Pharmaceutically acceptable carriers are biologically compatible vehicles which are suitable for administration to an animal e.g., physiological saline. A therapeutically effective amount is an amount of the nucleic acid of the invention which is capable of producing a medically desirable result in a patient. As is well known in the medical arts, dosage for any given patient depends upon many factors, including the patient's size, body surface area, age, the particular compound to be administered, sex, time and route of administration, general health, and other drugs being administered concurrently. Dosages will vary, but a preferred dosage for intravenous administration of a nucleic acid is from approximately 10^6 to 10^{22} copies of the nucleic acid molecule.

Determination of optimal dosage is well within the abilities of a pharmacologist of ordinary skill.

Example 7: AIB1 Knockout and Overexpression Mouse Mutants

Mutants organism that underexpress or overexpress AIB1 are useful for research. Such mutants allow insight into the physiological and/or pathological role of AIB1 in a healthy and/or pathological organism. These mutants are said to be "genetically engineered," meaning that information in the form of nucleotides has been transferred into the mutant's genome at a location, or in a combination, in which it would not normally exist. Nucleotides transferred in this way are said to be "non-native." For example, a WAP promoter inserted upstream of a native AIB1 gene would be non-native. An extra copy of a mouse AIB1 gene present on a plasmid and transformed into a mouse cell would be non-native. Mutants may be, for example, produced from mammals, such as mice, that either overexpress AIB1 or underexpress AIB1 or that do not express AIB1 at all. Overexpression mutants are made by increasing the number of AIB1 genes in the organism, or by introducing an AIB1 gene into the organism under the control of a constitutive or inducible or viral promoter such as the mouse mammary tumor virus (MMTV) promoter or the whey acidic protein (WAP) promoter or the metallothionein promoter. Mutants that underexpress AIB1 may be made by using an inducible or repressible promoter, or by deleting the AIB1 gene, or by destroying or limiting the function of the AIB1 gene, for instance by disrupting the gene by transposon insertion.

Anti-sense genes may be engineered into the organism, under a constitutive or inducible promoter, to decrease or prevent AIB1 expression. A gene is said to be "functionally deleted" when genetic engineering has been used to negate or reduce gene expression to negligible levels. When a mutant is referred to in this application as having the AIB1 gene altered or functionally deleted, this reference refers to the AIB1 gene and to any ortholog of this gene, for instance "a transgenic animal wherein at least one AIB1 gene has been functionally deleted" would encompass the mouse ortholog of the AIB1 gene, pCIP. When a mutant is referred to as having "more than the normal copy number" of a gene, this means that it has more than the usual number of genes found in the wild-type organism, eg: in the diploid mouse or human.

A mutant mouse overexpressing AIB1 may be made by constructing a plasmid having the AIB1 gene driven by a promoter, such as the mouse mammary tumor virus (MMTV) promoter or the whey acidic protein (WAP) promoter. This plasmid may be introduced into mouse oocytes by microinjection. The oocytes are implanted into pseudopregnant females, and the litters are assayed for insertion of the transgene. Multiple strains containing the transgene are then available for study.

WAP is quite specific for mammary gland expression during lactation, and MMTV is expressed in a variety of tissues including mammary gland, salivary gland and lymphoid tissues.

Many other promoters might be used to achieve various paterns of expression, e.g., the metallothionein promoter.

An inducible system may be created in which AIB1 is driven by a promoter regulated by an agent which can be fed to the mouse such as tetracycline. Such techniques are well known in the art.

A mutant knockout mouse from which the AIB1 (also called pCIP) gene is deleted was made by removing coding regions of the AIB1 gene from mouse embryonic stem cells. Fig. 5 shows the intron/exon structure for pCIP. Using this table, mutations can be targeted to coding sequences, avoiding silent mutations caused by deletion of non-coding sequences. (Fig. 6 shows the intron/exon structure for the human AIB1 gene). These cells were microinjected into mouse embryos leading to the deletion of the mouse AIB1 gene in the germ line of a transgenic mouse. The mouse AIB1 gene was mapped and isolated by the following method: The primers AIB/mEST F1 (5'-TCCTTTTCCCAGCAGCAGTTTG-3'; SEQ.I.D. 10) and AIB1/mEST R1 (5'-ATGCCAGACATGGGCATGGG-3' SEQ.I.D.11) were used to screen a mouse Bacterial Artificial Chromosome (BAC) library and to isolate a mouse BAC (designated 195H10). This BAC was assigned to mouse chromosome 2 by fluorescence in situ hybridization (FISH). This region is the mouse equivalent of the portion of human chromosome 20 which carries AIB1.

To map the structure of the gene, first the structure of the human AIB1 gene was determined by polymerase chain reaction of a human genomic DNA clone containing AIB1 using standard methods (Genomics 1995 Jan 20;25(2):501-506) and then the sequences of the intron exon boundaries were determined (Fig.4). Based on this information, the corresponding regions of the mouse BAC were sequenced. The structure of the mouse gene corresponds closely to that of the human gene (Fig. 4). This information localizes the coding regions of the mouse AIB1 gene so that a targeting vector can be constructed to remove these regions from mouse embryonic stem cells. These cells can be then injected into mouse embryos leading to deletion of the mouse AIB1 gene in the germ line of a transgenic mouse. The methods of creating deletion mutations by using a targeting vector have been described in Cell (Thomas and Capecch, Cell 51(3):503-512, 1987).

References and patents referred to herein are incorporated by reference.

The above examples are provided by way of illustration only and are in no way intended to limit the scope of the invention. One of skill in the art will see that the invention may be modified in various ways without departing from the spirit or principle of the invention. We claim all such modifications.

Sequence Listing

(1) GENERAL INFORMATION

(i) APPLICANT: Meltzer and Trent

(ii) TITLE OF INVENTION: AIB1, A NOVEL RECEPTOR CO-ACTIVATOR AMPLIFIED IN CANCER

(iii) NUMBER OF SEQUENCES: 12

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Klarquist Sparkman Campbell Leigh & Whinston, LLP
(B) STREET: One World Trade Center
121 S.W. Salmon Street, Suite 1600
(C) CITY: Portland
(D) STATE: Oregon
(E) COUNTRY: United States of America
(F) ZIP: 97204-2988

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Disk, 3-1/2 inch
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: Widows NT
(D) SOFTWARE: WordPerfect 7.0 & ASCII

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: William D. Noonan, M.D.
(B) REGISTRATION NUMBER: 30,878
(C) REFERENCE/DOCKET NUMBER: 4239-49944

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (503) 226-7391
(B) TELEFAX: (503) 228-9446

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6837 nucleotides; 1419 amino acid residues
(B) TYPE: Human DNA & Amino Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
CG  GCG GCG GCT GCG GCT TAG TCG GTG GCG GCC GGC GGC GGC TGC GGG CTG AGC GGC
1               5                  10                  15
GAG TTT CCG ATT TAA AGC TGA GCT GCG AGG AAA ATG GCG GCG GGA GGA TCA AAA TAC
20                  25                  30                  35
```

```
TTG CTG GAT GGT GGA CTC AGA GAC CAA TAA AAA TAA ACT GCT TGA ACA TCC TTT GAC
    40                  45                  50                  55
TGG TTA GCC AGT TGC TGA TGT ATA TTC AAG ATG AGT GGA TTA GGA GAA AAC TTG GAT
                                        Met Ser Gly Leu Gly Glu Asn Leu Asp
        60                  65                  70                  75
CCA CTG GCC AGT GAT TCA CGA AAA CGC AAA TTG CCA TGT GAT ACT CCA GGA CAA GGT
Pro Leu Ala Ser Asp Ser Arg Lys Arg Lys Leu Pro Cys Asp Thr Pro Gly Gln Gly
                80                  85                  90                  95
CTT ACC TGC AGT GGT GAA AAA CGG AGA CGG GAG CAG GAA AGT AAA TAT ATT GAA GAA
Leu Thr Cys Ser Gly Glu Lys Arg Arg Arg Glu Gln Glu Ser Lys Tyr Ile Glu Glu
            100                 105                 110
TTG GCT GAG CTG ATA TCT GCC AAT CTT AGT GAT ATT GAC AAT TTC AAT GTC AAA CCA
Leu Ala Glu Leu Ile Ser Ala Asn Leu Ser Asp Ile Asp Asn Phe Asn Val Lys Pro
115                 120                 125                 130
GAT AAA TGT GCG ATT TTA AAG GAA ACA GTA AGA CAG ATA CGT CAA ATA AAA GAG CAA
Asp Lys Cys Ala Ile Leu Lys Glu Thr Val Arg Gln Ile Arg Gln Ile Lys Glu Gln
    135                 140                 145                 150
GGA AAA ACT ATT TCC AAT GAT GAT GAT GTT CAA AAA GCC GAT GTA TCT TCT ACA GGG
Gly Lys Thr Ile Ser Asn Asp Asp Asp Val Gln Lys Ala Asp Val Ser Ser Thr Gly
        155                 160                 165                 170
CAG GGA GTT ATT GAT AAA GAC TCC TTA GGA CCG CTT TTA CTT CAG GCA TTG GAT GGT
Gln Gly Val Ile Asp Lys Asp Ser Leu Gly Pro Leu Leu Leu Gln Ala Leu Asp Gly
            175                 180                 185                 190
TTC CTA TTT GTG GTG AAT CGA GAC GGA AAC ATT GTA TTT GTA TCA GAA AAT GTC ACA
Phe Leu Phe Val Val Asn Arg Asp Gly Asn Ile Val Phe Val Ser Glu Asn Val Thr
            195                 200                 205
CAA TAC CTG CAA TAT AAG CAA GAG GAC CTG GTT AAC ACA AGT GTT TAC AAT ATC TTA
Gln Tyr Leu Gln Tyr Lys Gln Glu Asp Leu Val Asn Thr Ser Val Tyr Asn Ile Leu
210                 215                 220                 225
CAT GAA GAA GAC AGA AAG GAT TTT CTT AAG AAT TTA CCA AAA TCT ACA GTT AAT GGA
His Glu Glu Asp Arg Lys Asp Phe Leu Lys Asn Leu Pro Lys Ser Thr Val Asn Gly
    230                 235                 240                 245
GTT TCC TGG ACA AAT GAG ACC CAA AGA CAA AAA AGC CAT ACA TTT AAT TGC CGT ATG
Val Ser Trp Thr Asn Glu Thr Gln Arg Gln Lys Ser His Thr Phe Asn Cys Arg Met
        250                 255                 260                 265
TTG ATG AAA ACA CCA CAT GAT ATT CTG GAA GAC ATA AAC GCC AGT CCT GAA ATG CGC
Leu Met Lys Thr Pro His Asp Ile Leu Glu Asp Ile Asn Ala Ser Pro Glu Met Arg
            270                 275                 280                 285
CAG AGA TAT GAA ACA ATG CAG TGC TTT GCC CTG TCT CAG CCA CGA GCT ATG ATG GAG
Gln Arg Tyr Glu Thr Met Gln Cys Phe Ala Leu Ser Gln Pro Arg Ala Met Met Glu
            290                 295                 300
GAA GGG GAA GAT TTG CAA TCT TGT ATG ATC TGT GTG GCA CGC CGC ATT ACT ACA GGA
Glu Gly Glu Asp Leu Gln Ser Cys Met Ile Cys Val Ala Arg Arg Ile Thr Thr Gly
305                 310                 315                 320
GAA AGA ACA TTT CCA TCA AAC CCT GAG AGC TTT ATT ACC AGA CAT GAT CTT TCA GGA
Glu Arg Thr Phe Pro Ser Asn Pro Glu Ser Phe Ile Thr Arg His Asp Leu Ser Gly
    325                 330                 335                 340
AAG GTT GTC AAT ATA GAT ACA AAT TCA CTG AGA TCC TCC ATG AGG CCT GGC TTT GAA
Lys Val Val Asn Ile Asp Thr Asn Ser Leu Arg Ser Ser Met Arg Pro Gly Phe Glu
        345                 350                 355                 360
GAT ATA ATC CGA AGG TGT ATT CAG AGA TTT TTT AGT CTA AAT GAT GGG CAG TCA TGG
Asp Ile Ile Arg Arg Cys Ile Gln Arg Phe Phe Ser Leu Asn Asp Gly Gln Ser Trp
            365                 370                 375                 380
TCC CAG AAA CGT CAC TAT CAA GAA GCT TAT CTT AAT GGC CAT GCA GAA ACC CCA GTA
Ser Gln Lys Arg His Tyr Gln Glu Ala Tyr Leu Asn Gly His Ala Glu Thr Pro Val
            385                 390                 395
TAT CGA TTC TCG TTG GCT GAT GGA ACT ATA GTG ACT GCA CAG ACA AAA AGC AAA CTC
Tyr Arg Phe Ser Leu Ala Asp Gly Thr Ile Val Thr Ala Gln Thr Lys Ser Lys Leu
400                 405                 410                 415

TTC CGA AAT CCT GTA ACA AAT GAT CGA CAT GGC TTT GTC TCA ACC CAC TTC CTT CAG
Phe Arg Asn Pro Val Thr Asn Asp Arg His Gly Phe Val Ser Thr His Phe Leu Gln
    420                 425                 430                 435
AGA GAA CAG AAT GGA TAT AGA CCA AAC CCA AAT CCT GTT GGA CAA GGG ATT AGA CCA
Arg Glu Gln Asn Gly Tyr Arg Pro Asn Pro Asn Pro Val Gly Gln Gly Ile Arg Pro
        440                 445                 450                 455
CCT ATG GCT GGA TGC AAC AGT TCG GTA GGC GGC ATG AGT ATG TCG CCA AAC CAA GGC
Pro Met Ala Gly Cys Asn Ser Ser Val Gly Gly Met Ser Met Ser Pro Asn Gln Gly
            460                 465                 470                 475
TTA CAG ATG CCG AGC AGC AGG GCC TAT GGC TTG GCA GAC CCT AGC ACC ACA GGG CAG
Leu Gln Met Pro Ser Ser Arg Ala Tyr Gly Leu Ala Asp Pro Ser Thr Thr Gly Gln
```

```
                                480                 485                 490
ATG AGT GGA GCT AGG TAT GGG GGT TCC AGT AAC ATA GCT TCA TTG ACC CCT GGG CCA
Met Ser Gly Ala Arg Tyr Gly Gly Ser Ser Asn Ile Ala Ser Leu Thr Pro Gly Pro
495                 500                 505                 510
GGC ATG CAA TCA CCA TCT TCC TAC CAG AAC AAC AAC TAT GGG CTC AAC ATG AGT AGC
Gly Met Gln Ser Pro Ser Ser Tyr Gln Asn Asn Asn Tyr Gly Leu Asn Met Ser Ser
    515                 520                 525                 530
CCC CCA CAT GGG AGT CCT GGT CTT GCC CCA AAC CAG CAG AAT ATC ATG ATT TCT CCT
Pro Pro His Gly Ser Pro Gly Leu Ala Pro Asn Gln Gln Asn Ile Met Ile Ser Pro
        535                 540                 545                 550
CGT AAT CGT GGG AGT CCA AAG ATA GCC TCA CAT CAG TTT TCT CCT GTT GCA GGT GTG
Arg Asn Arg Gly Ser Pro Lys Ile Ala Ser His Gln Phe Ser Pro Val Ala Gly Val
            555                 560                 565                 570
CAC TCT CCC ATG GCA TCT TCT GGC AAT ACT GGG AAC CAC AGC TTT TCC AGC AGC TCT
His Ser Pro Met Ala Ser Ser Gly Asn Thr Gly Asn His Ser Phe Ser Ser Ser Ser
                575                 580                 585
CTC AGT GCC CTG CAA GCC ATC AGT GAA GGT GTG GGG ACT TCC CTT TTA TCT ACT CTG
Leu Ser Ala Leu Gln Ala Ile Ser Glu Gly Val Gly Thr Ser Leu Leu Ser Thr Leu
590                 595                 600                 605
TCA TCA CCA GGC CCC AAA TTG GAT AAC TCT CCC AAT ATG AAT ATT ACC CAA CCA AGT
Ser Ser Pro Gly Pro Lys Leu Asp Asn Ser Pro Asn Met Asn Ile Thr Gln Pro Ser
    610                 615                 620                 625
AAA GTA AGC AAT CAG GAT TCC AAG AGT CCT CTG GGC TTT TAT TGC GAC CAA AAT CCA
Lys Val Ser Asn Gln Asp Ser Lys Ser Pro Leu Gly Phe Tyr Cys Asp Gln Asn Pro
        630                 635                 640                 645
GTG GAG AGT TCA ATG TGT CAG TCA AAT AGC AGA GAT CAC CTC AGT GAC AAA GAA AGT
Val Glu Ser Ser Met Cys Gln Ser Asn Ser Arg Asp His Leu Ser Asp Lys Glu Ser
            650                 655                 660                 665
AAG GAG AGC AGT GTT GAG GGG GCA GAG AAT CAA AGG GGT CCT TTG GAA AGC AAA GGT
Lys Glu Ser Ser Val Glu Gly Ala Glu Asn Gln Arg Gly Pro Leu Glu Ser Lys Gly
                670                 675                 680
CAT AAA AAA TTA CTG CAG TTA CTT ACC TGT TCT TCT GAT GAC CGG GGT CAT TCC TCC
His Lys Lys Leu Leu Gln Leu Leu Thr Cys Ser Ser Asp Asp Arg Gly His Ser Ser
685                 690                 695                 700
TTG ACC AAC TCC CCC CTA GAT TCA AGT TGT AAA GAA TCT TCT GTT AGT GTC ACC AGC
Leu Thr Asn Ser Pro Leu Asp Ser Ser Cys Lys Glu Ser Ser Val Ser Val Thr Ser
    705                 710                 715                 720
CCC TCT GGA GTC TCC TCC TCT ACA TCT GGA GGA GTA TCC TCT ACA TCC AAT ATG CAT
Pro Ser Gly Val Ser Ser Ser Thr Ser Gly Gly Val Ser Ser Thr Ser Asn Met His
        725                 730                 735                 740
GGG TCA CTG TTA CAA GAG AAG CAC CGG ATT TTG CAC AAG TTG CTG CAG AAT GGG AAT
Gly Ser Leu Leu Gln Glu Lys His Arg Ile Leu His Lys Leu Leu Gln Asn Gly Asn
            745                 750                 755                 760
TCA CCA GCT GAG GTA GCC AAG ATT ACT GCA GAA GCC ACT GGG AAA GAC ACC AGC AGT
Ser Pro Ala Glu Val Ala Lys Ile Thr Ala Glu Ala Thr Gly Lys Asp Thr Ser Ser
                765                 770                 775
ATA ACT TCT TGT GGG GAC GGA AAT GTT GTC AAG CAG GAG CAG CTA AGT CCT AAG AAG
Ile Thr Ser Cys Gly Asp Gly Asn Val Val Lys Gln Glu Gln Leu Ser Pro Lys Lys
780                 785                 790                 795
AAG GAG AAT AAT GCA CTT CTT AGA TAC CTG CTG GAC AGG GAT GAT CCT AGT GAT GCA
Lys Glu Asn Asn Ala Leu Leu Arg Tyr Leu Leu Asp Arg Asp Asp Pro Ser Asp Ala
    800                 805                 810                 815

CTC TCT AAA GAA CTA CAG CCC CAA GTG GAA GGA GTG GAT AAT AAA ATG AGT CAG TGC
Leu Ser Lys Glu Leu Gln Pro Gln Val Glu Gly Val Asp Asn Lys Met Ser Gln Cys
        820                 825                 830                 835
ACC AGC TCC ACC ATT CCT AGC TCA AGT CAA GAG AAA GAC CCT AAA ATT AAG ACA GAG
Thr Ser Ser Thr Ile Pro Ser Ser Ser Gln Glu Lys Asp Pro Lys Ile Lys Thr Glu
            840                 845                 850                 855
ACA AGT GAA GAG GGA TCT GGA GAC TTG GAT AAT CTA GAT GCT ATT CTT GGT GAT CTG
Thr Ser Glu Glu Gly Ser Gly Asp Leu Asp Asn Leu Asp Ala Ile Leu Gly Asp Leu
                860                 865                 870
ACT AGT TCT GAC TTT TAC AAT AAT TCC ATA TCC TCA AAT GGT AGT CAT CTG GGG ACT
Thr Ser Ser Asp Phe Tyr Asn Asn Ser Ile Ser Ser Asn Gly Ser His Leu Gly Thr
875                 880                 885                 890
AAG CAA CAG GTG TTT CAA GGA ACT AAT TCT CTG GGT TTG AAA AGT TCA CAG TCT GTG
Lys Gln Gln Val Phe Gln Gly Thr Asn Ser Leu Gly Leu Lys Ser Ser Gln Ser Val
    895                 900                 905                 910
CAG TCT ATT CGT CCT CCA TAT AAC CGA GCA GTG TCT CTG GAT AGC CCT GTT TCT GTT
Gln Ser Ile Arg Pro Pro Tyr Asn Arg Ala Val Ser Leu Asp Ser Pro Val Ser Val
        915                 920                 925                 930
```

```
GGC TCA AGT CCT CCA GTA AAA AAT ATC AGT GCT TTC CCC ATG TTA CCA AAG CAA CCC
Gly Ser Ser Pro Pro Val Lys Asn Ile Ser Ala Phe Pro Met Leu Pro Lys Gln Pro
            935                 940                 945                 950
ATG TTG GGT GGG AAT CCA AGA ATG ATG GAT AGT CAG GAA AAT TAT GGC TCA AGT ATG
Met Leu Gly Gly Asn Pro Arg Met Met Asp Ser Gln Glu Asn Tyr Gly Ser Ser Met
                955                 960                 965
GGT GGG CCA AAC CGA AAT GTG ACT GTG ACT CAG ACT CCT TCC TCA GGA GAC TGG GGC
Gly Gly Pro Asn Arg Asn Val Thr Val Thr Gln Thr Pro Ser Ser Gly Asp Trp Gly
970                 975                 980                 985
TTA CCA AAC TCA AAG GCC GGC AGA ATG GAA CCT ATG AAT TCA AAC TCC ATG GGA AGA
Leu Pro Asn Ser Lys Ala Gly Arg Met Glu Pro Met Asn Ser Asn Ser Met Gly Arg
    990                 995                1000                1005
CCA GGA GGA GAT TAT AAT ACT TCT TTA CCC AGA CCT GCA CTG GGT GGC TCT ATT CCC
Pro Gly Gly Asp Tyr Asn Thr Ser Leu Pro Arg Pro Ala Leu Gly Gly Ser Ile Pro
        1010                1015                1020                1025
ACA TTG CCT CTT CGG TCT AAT AGC ATA CCA GGT GCG AGA CCA GTA TTG CAA CAG CAG
Thr Leu Pro Leu Arg Ser Asn Ser Ile Pro Gly Ala Arg Pro Val Leu Gln Gln Gln
            1030                1035                1040                1045
CAG CAG ATG CTT CAA ATG AGG CCT GGT GAA ATC CCC ATG GGA ATG GGG GCT AAT CCC
Gln Gln Met Leu Gln Met Arg Pro Gly Glu Ile Pro Met Gly Met Gly Ala Asn Pro
                1050                1055                1060
TAT GGC CAA GCA GCA GCA TCT AAC CAA CTG GGT TCC TGG CCC GAT GGC ATG TTG TCC
Tyr Gly Gln Ala Ala Ala Ser Asn Gln Leu Gly Ser Trp Pro Asp Gly Met Leu Ser
1065                1070                1075                1080
ATG GAA CAA GTT TCT CAT GGC ACT CAA AAT AGG CCT CTT CTT AGG AAT TCC CTG GAT
Met Glu Gln Val Ser His Gly Thr Gln Asn Arg Pro Leu Leu Arg Asn Ser Leu Asp
    1085                1090                1095                1100
GAT CTT GTT GGG CCA CCT TCC AAC CTG GAA GGC CAG AGT GAC GAA AGA GCA TTA TTG
Asp Leu Val Gly Pro Pro Ser Asn Leu Glu Gly Gln Ser Asp Glu Arg Ala Leu Leu
        1105                1110                1115                1120
GAC CAG CTG CAC ACT CTT CTC AGC AAC ACA GAT GCC ACA GGC CTG GAA GAA ATT GAC
Asp Gln Leu His Thr Leu Leu Ser Asn Thr Asp Ala Thr Gly Leu Glu Glu Ile Asp
            1125                1130                1135                1140
AGA GCT TTG GGC ATT CCT GAA CTT GTC AAT CAG GGA CAG GCA TTA GAG CCC AAA CAG
Arg Ala Leu Gly Ile Pro Glu Leu Val Asn Gln Gly Gln Ala Leu Glu Pro Lys Gln
                1145                1150                1155
GAT GCT TTC CAA GGC CAA GAA GCA GCA GTA ATG ATG GAT CAG AAG GCA GGA TTA TAT
Asp Ala Phe Gln Gly Gln Glu Ala Ala Val Met Met Asp Gln Lys Ala Gly Leu Tyr
1160                1165                1170                1175
GGA CAG ACA TAC CCA GCA CAG GGG CCT CCA ATG CAA GGA GGC TTT CAT CTT CAG GGA
Gly Gln Thr Tyr Pro Ala Gln Gly Pro Pro Met Gln Gly Gly Phe His Leu Gln Gly
    1180                1185                1190                1195
CAA TCA CCA TCT TTT AAC TCT ATG ATG AAT CAG ATG AAC CAG CAA GGC AAT TTT CCT
Gln Ser Pro Ser Phe Asn Ser Met Met Asn Gln Met Asn Gln Gln Gly Asn Phe Pro
        1200                1205                1210                1215

CTC CAA GGA ATG CAC CCA CGA GCC AAC ATC ATG AGA CCC CGG ACA AAC ACC CCC AAG
Leu Gln Gly Met His Pro Arg Ala Asn Ile Met Arg Pro Arg Thr Asn Thr Pro Lys
            1220                1225                1230                1235
CAA CTT AGA ATG CAG CTT CAG CAG AGG CTG CAG GGC CAG CAG TTT TTG AAT CAG AGC
Gln Leu Arg Met Gln Leu Gln Gln Arg Leu Gln Gly Gln Gln Phe Leu Asn Gln Ser
                1240                1245                1250
CGA CAG GCA CTT GAA TTG AAA ATG GAA AAC CCT ACT GCT GGT GGT GCT GCG GTG ATG
Arg Gln Ala Leu Glu Leu Lys Met Glu Asn Pro Thr Ala Gly Gly Ala Ala Val Met
1255                1260                1265                1270
AGG CCT ATG ATG CAG CCC CAG CAG GGT TTT CTT AAT GCT CAA ATG GTC GCC CAA CGC
Arg Pro Met Met Gln Pro Gln Gln Gly Phe Leu Asn Ala Gln Met Val Ala Gln Arg
    1275                1280                1285                1290
AGC AGA GAG CTG CTA AGT CAT CAC TTC CGA CAA CAG AGG GTG GCT ATG ATG ATG CAG
Ser Arg Glu Leu Leu Ser His His Phe Arg Gln Gln Arg Val Ala Met Met Met Gln
        1295                1300                1305                1310
CAG CAG CAG CAG CAG CAA CAG CAG CAG CAG CAG CAG CAG CAG CAG CAA CAG CAA CAG
Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln
            1315                1320                1325                1330
CAA CAG CAA CAG CAG CAA CAG CAG CAA ACC CAG GCC TTC AGC CCA CCT CCT AAT GTG
Gln Gln Gln Gln Gln Gln Gln Gln Thr Gln Ala Phe Ser Pro Pro Pro Asn Val
                1335                1340                1345
ACT GCT TCC CCC AGC ATG GAT GGG CTT TTG GCA GGA CCC ACA ATG CCA CAA GCT CCT
Thr Ala Ser Pro Ser Met Asp Gly Leu Leu Ala Gly Pro Thr Met Pro Gln Ala Pro
1350                1355                1360                1365
CCG CAA CAG TTT CCA TAT CAA CCA AAT TAT GGA ATG GGA CAA CAA CCA GAT CCA GCC
```

```
Pro Gln Gln Phe Pro Tyr Gln Pro Asn Tyr Gly Met Gly Gln Gln Pro Asp Pro Ala
    1370                1375                1380                1385
TTT GGT CGA GTG TCT AGT CCT CCC AAT GCA ATG ATG TCG TCA AGA ATG GGT CCC TCC
Phe Gly Arg Val Ser Ser Pro Pro Asn Ala Met Met Ser Ser Arg Met Gly Pro Ser
        1390                1395                1400                1405
CAG AAT CCC ATG ATG CAA CAC CCG CAG GCT GCA TCC ATC TAT CAG TCC TCA GAA ATG
Gln Asn Pro Met Met Gln His Pro Gln Ala Ala Ser Ile Tyr Gln Ser Ser Glu Met
            1410                1415                1420                1425
AAG GGC TGG CCA TCA GGA AAT TTG GCC AGG AAC AGC TCC TTT TCC CAG CAG CAG TTT
Lys Gly Trp Pro Ser Gly Asn Leu Ala Arg Asn Ser Ser Phe Ser Gln Gln Gln Phe
                1430                1435                1440
GCC CAC CAG GGG AAT CCT GCA GTG TAT AGT ATG GTG CAC ATG AAT GGC AGC AGT GGT
Ala His Gln Gly Asn Pro Ala Val Tyr Ser Met Val His Met Asn Gly Ser Ser Gly
1445                1450                1455                1460
CAC ATG GGA CAG ATG AAC ATG AAC CCC ATG CCC ATG TCT GGC ATG CCT ATG GGT CCT
His Met Gly Gln Met Asn Met Asn Pro Met Pro Met Ser Gly Met Pro Met Gly Pro
    1465                1470                1475                1480
GAT CAG AAA TAC TGC TGA CAT CTC TGC ACC AGG ACC TCT TAA GGA AAC CAC TGT ACA
Asp Gln Lys Tyr Cys ***
        1485                1490                1495                1500
AAT GAC ACT GCA CTA GGA TTA TTG GGA AGG AAT CAT TGT TCC AGG CAT CCA TCT TGG
            1505                1510                1515                1520
AAG AAA GGA CCA GCT TTG AGC TCC ATC AAG GGT ATT TTA AGT GAT GTC ATT TGA GCA
                1525                1530                1535
GGA CTG GAT TTT AAG CCG AAG GGC AAT ATC TAC GTG TTT TTC CCC CCT CCT TCT GCT
1540                1545                1550                1555
GTG TAT CAT GGT GTT CAA AAC AGA AAT GTT TTT TGG CAT TCC ACC TCC TAG GGA TAT
    1560                1565                1570                1575
AAT TCT GGA GAC ATG GAG TGT TAC TGA TCA TAA AAC TTT TGT GTC ACT TTT TTC TGC
        1580                1585                1590                1595
CTT GCT AGC CAA AAT CTC TTA AAT ACA CGT AGG TGG GCC AGA GAA CAT TGG AAG AAT
            1600                1605                1610                1615
CAA GAG AGA TTA GAA TAT CTG GTT TCT CTA GTT GCA GTA TTG GAC AAA GAG CAT AGT
                1620                1625                1630
CCC AGC CTT CAG GTG TAG TAG TTC TGT GTT GAC CCT TTG TCC AGT GGA ATT GGT GAT
1635                1640                1645                1650
TCT GAA TTG TCC TTT ACT AAT GGT GTT GAG TTG CTC TGT CCC TAT TAT TTG CCC TAG
    1655                1660                1665                1670

GCT TTC TCC TAA TGA AGG TTT TCA TTT GCC ATT CAT GTC CTG TAA TAC TTC ACC TCC
        1675                1680                1685                1690
AGG AAC TGT CAT GGA TGT CCA AAT GGC TTT GCA GAA AGG AAA TGA GAT GAC AGT ATT
            1695                1700                1705                1710
TAA TCG CAG CAG TAG CAA ACT TTT CAC ATG CTA ATG TGC AGC TGA GTG CAC TTT ATT
                1715                1720                1725
TAA AAA GAA TGG ATA AAT GCA ATA TTC TTG AGG TCT TGA GGG AAT AGT GAA ACA CAT
1730                1735                1740                1745
TCC TGG TTT TTG CCT ACA CTT ACG TGT TAG ACA AGA ACT ATG ATT TTT TTT TTA AAG
    1750                1755                1760                1765
TAC TGG TGT CAC CCT TTG CCT ATA TGG TAG AGC AAT AAT GCT TTT TAA AAA TAA ACT
        1770                1775                1780                1785
TCT GAA AAC CCA AGG CCA GGT ACT GCA TTC TGA ATC AGA ATC TCG CAG TGT TTC TGT
            1790                1795                1800                1805
GAA TAG ATT TTT TTG TAA ATA TGA CCT TTA AGA TAT TGT ATT ATG TAA AAT ATG TAT
                1810                1815                1820
ATA CCT TTT TTT GTA GGT CAC AAC AAC TCA TTT TTA CAG AGT TTG TGA AGC TAA ATA
1825                1830                1835                1840
TTT AAC ATT GTT GAT TTC AGT AAG CTG TGT GGT GAG GCT ACC AGT GGA AGA GAC ATC
    1845                1850                1855                1860
CCT TGA CTT TTG TGG CCT GGG GGA GGG GTA GTG CTC CAC AGC TTT TCC TTC CCC ACC
        1865                1870                1875                1880
CCC CAG CCT TAG ATG CCT CGC TCT TTT CAA TCT CTT AAT CTA AAT GCT TTT TAA AGA
            1885                1890                1895                1900
GAT TAT TTG TTT AGA TGT AGG CAT TTT AAT TTT TTA AAA ATT CCT CTA CCA GAA CTA
                1905                1910                1915
AGC ACT TTG TTA ATT TGG GGG GAA AGA ATA GAT ATG GGG AAA TAA ACT TAA AAA AAA
1920                1925                1930                1935
ATC AGG AAT TTA AAA AAA CGA GCA ATT TGA AGA GAA TCT TTT GGA TTT TAA GCA GTC
    1940                1945                1950                1955
CGA AAT AAT AGC AAT TCA TGG GCT GTG TGT GTG TGT GTA TGT GTG TGT GTG TGT GTG
        1960                1965                1970                1975
```

```
TAT GTT TAA TTA TGT TAC CTT TTC ATC CCC TTT AGG AGC GTT TTC AGA TTT TGG TTG
            1980                1985                1990                1995
CTA AGA CCT GAA TCC CAT ATT GAG ATC TCG AGT AGA ATC CTT GGT GTG GTT TCT GGT
                  2000                2005                2010
GTC TGC TCA GCT GTC CCC TCA TTC TAC TAA TGT GAT GCT TTC ATT ATG TCC CTG TGG
2015                2020                2025                2030
ATT AGA ATA GTG TCA GTT ATT TCT TAA GTA ACT CAG TAC CCA GAA CAG CCA GTT TTA
    2035                2040                2045                2050
CTG TGA TTC AGA GCC ACA GTC TAA CTG AGC ACC TTT TAA ACC CCT CCC TCT TCT GCC
          2055                2060                2065                2070
CCC TAC CAC TTT TCT GCT GTT GCC TCT CTT TGA CAC CTG TTT TAG TCA GTT GGG AGG
                  2075                2080                2085                2090
AAG GGA AAA ATC AAG TTT AAT TCC CTT TAT CTG GGT TAA TTC ATT TGG TTC AAA TAG
                  2095                2100                2105
TTG ACG GAA TTG GGT TTC TGA ATG TCT GTG AAT TTC AGA GGT CTC TGC TAG CCT TGG
2110                2115                2120                2125
TAT CAT TTT CTA GCA ATA ACT GAG AGC CAG TTA ATT TTA AGA ATT TCA CAC ATT TAG
    2130                2135                2140                2145
CCA ATC TTT CTA GAT GTC TCT GAA GGT AAG ATC ATT TAA TAT CTT TGA TAT GCT TAC
          2150                2155                2160                2165
GAG TAA GTG AAT CCT GAT TAT TTC CAG ACC CAC CAC CAG AGT GGA TCT TAT TTT CAA
                  2170                2175                2180                2185
AGC AGT ATA GAC AAT TAT GAG TTT GCC CTC TTT CCC CTA CCA AGT TCA AAA TAT ATC
                  2190                2195                2200
TAA GAA AGA TTG TAA ATC CGA AAA CTT CCA TTG TAG TGG CCT GTG CTT TTC AGA TAG
2205                2210                2215                2220
TAT ACT CTC CTG TTT GGA GAC AGA GGA AGA ACC AGG TCA GTC TGT CTC TTT TTC AGC
    2225                2230                2235                2240
TCA ATT GTA TCT GAC CCT TCT TTA AGT TAT GTG TGT GGG GAG AAA TAG AAT GGT GCT
          2245                2250                2255                2260
CTT ATC TTT CTT GAC TTT AAA AAA ATT ATT AAA AAC AAA AAA AAA AAA AAA AA
                  2265                2270                2275
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 186

(B) TYPE: amino acid

(C) STRANDEDNESS: Single

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Leu Leu Gln Ala Leu Asp Gly Phe Leu Phe Val Val Asn Arg Asp Gly Asn Ile Val
1               5                  10                  15
Phe Val Ser Glu Asn Val Thr Gln Tyr Leu Gln Tyr Lys Gln Glu Asp Leu Val Asn
20                  25                  30                  35
Thr Ser Val Tyr Asn Ile Leu His Glu Glu Asp Arg Lys Asp Phe Leu Lys Asn Leu
            40                  45                  50                  55
Pro Lys Ser Thr Val Asn Gly Val Ser Trp Thr Asn Glu Thr Gln Arg Gln Lys Ser
                60                  65                  70                  75
His Thr Phe Asn Cys Arg Met Leu Met Lys Thr Pro His Asp Ile Leu Glu Asp Ile
                80                  85                  90
Asn Ala Ser Pro Glu Met Arg Gln Arg Tyr Glu Thr Met Gln Cys Phe Ala Leu Ser
95                  100                 105                 110
Gln Pro Arg Ala Met Met Glu Glu Gly Glu Asp Leu Gln Ser Cys Met Ile Cys Val
    115                 120                 125                 130
Ala Arg Arg Ile Thr Thr Gly Glu Arg Thr Phe Pro Ser Asn Pro Glu Ser Phe Ile
            135                 140                 145                 150
Thr Arg His Asp Leu Ser Gly Lys Val Val Asn Ile Asp Thr Asn Ser Leu Arg Ser
                155                 160                 165                 170
Ser Met Arg Pro Gly Phe Glu Asp Ile Ile Arg Arg Cys Ile Gln
                175                 180                 185
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 73

(B) TYPE: amino acid

(C) STRANDEDNESS: Single

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Arg Lys Arg Lys Leu Pro Cys Asp Thr Pro Gly Gln Gly Leu Thr Cys Ser Gly Glu
1               5                  10                  15
Lys Arg Arg Arg Glu Gln Glu Ser Lys Tyr Ile Glu Glu Leu Ala Glu Leu Ile Ser
20                  25                  130                 135
Ala Asn Leu Ser Asp Ile Asp Asn Phe Asn Val Lys Pro Asp Lys Cys Ala Ile Leu
    140                 145                 150                 155
Lys Glu Thr Val Arg Gln Ile Arg Gln Ile Lys Glu Gln Gly Lys Thr
        160                 165                 170
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1419

(B) TYPE: human amino acid of AIB1

(C) STRANDEDNESS: Single

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Ser Gly Leu Gly Glu Asn Leu Asp Pro Leu Ala Ser Asp Ser Arg Lys Arg Lys
1               5                  10                  15
Leu Pro Cys Asp Thr Pro Gly Gln Gly Leu Thr Cys Ser Gly Glu Lys Arg Arg Arg
20                  25                  30                  35
Glu Gln Glu Ser Lys Tyr Ile Glu Glu Leu Ala Glu Leu Ile Ser Ala Asn Leu Ser
    40                  45                  50                  55
Asp Ile Asp Asn Phe Asn Val Lys Pro Asp Lys Cys Ala Ile Leu Lys Glu Thr Val
        60                  65                  70                  75
Arg Gln Ile Arg Gln Ile Lys Glu Gln Gly Lys Thr Ile Ser Asn Asp Asp Asp Val
            80                  85                  90                  95
Gln Lys Ala Asp Val Ser Ser Thr Gly Gln Gly Val Ile Asp Lys Asp Ser Leu Gly
                100                 105                 110
Pro Leu Leu Leu Gln Ala Leu Asp Gly Phe Leu Phe Val Val Asn Arg Asp Gly Asn
115                 120                 125                 130
Ile Val Phe Val Ser Glu Asn Val Thr Gln Tyr Leu Gln Tyr Lys Gln Glu Asp Leu
    135                 140                 145                 150
Val Asn Thr Ser Val Tyr Asn Ile Leu His Glu Glu Asp Arg Lys Asp Phe Leu Lys
        155                 160                 165                 170
Asn Leu Pro Lys Ser Thr Val Asn Gly Val Ser Trp Thr Asn Glu Thr Gln Arg Gln
            175                 180                 185                 190
Lys Ser His Thr Phe Asn Cys Arg Met Leu Met Lys Thr Pro His Asp Ile Leu Glu
                195                 200                 205
Asp Ile Asn Ala Ser Pro Glu Met Arg Gln Arg Tyr Glu Thr Met Gln Cys Phe Ala
210                 215                 220                 225
Leu Ser Gln Pro Arg Ala Met Met Glu Glu Gly Glu Asp Leu Gln Ser Cys Met Ile
    230                 235                 240                 245
Cys Val Ala Arg Arg Ile Thr Thr Gly Glu Arg Thr Phe Pro Ser Asn Pro Glu Ser
        250                 255                 260                 265
Phe Ile Thr Arg His Asp Leu Ser Gly Lys Val Val Asn Ile Asp Thr Asn Ser Leu
            270                 275                 280                 285
Arg Ser Ser Met Arg Pro Gly Phe Glu Asp Ile Ile Arg Arg Cys Ile Gln Arg Phe
                290                 295                 300
Phe Ser Leu Asn Asp Gly Gln Ser Trp Ser Gln Lys Arg His Tyr Gln Glu Ala Tyr
305                 310                 315                 320
Leu Asn Gly His Ala Glu Thr Pro Val Tyr Arg Phe Ser Leu Ala Asp Gly Thr Ile
    325                 330                 335                 340
Val Thr Ala Gln Thr Lys Ser Lys Leu Phe Arg Asn Pro Val Thr Asn Asp Arg His
        345                 350                 355                 360
Gly Phe Val Ser Thr His Phe Leu Gln Arg Glu Gln Asn Gly Tyr Arg Pro Asn Pro
            365                 370                 375                 380
Asn Pro Val Gly Gln Gly Ile Arg Pro Pro Met Ala Gly Cys Asn Ser Ser Val Gly
                385                 390                 395
Gly Met Ser Met Ser Pro Asn Gln Gly Leu Gln Met Pro Ser Ser Arg Ala Tyr Gly
400                 405                 410                 415
Leu Ala Asp Pro Ser Thr Thr Gly Gln Met Ser Gly Ala Arg Tyr Gly Gly Ser Ser
    420                 425                 430                 435
Asn Ile Ala Ser Leu Thr Pro Gly Pro Gly Met Gln Ser Pro Ser Ser Tyr Gln Asn
        440                 445                 450                 455
```

```
Asn Asn Tyr Gly Leu Asn Met Ser Ser Pro Pro His Gly Ser Pro Gly Leu Ala Pro
            460                 465                 470                 475
Asn Gln Gln Asn Ile Met Ile Ser Pro Arg Asn Arg Gly Ser Pro Lys Ile Ala Ser
                480                 485                 490
His Gln Phe Ser Pro Val Ala Gly Val His Ser Pro Met Ala Ser Ser Gly Asn Thr
495                 500                 505                 510
Gly Asn His Ser Phe Ser Ser Ser Ser Leu Ser Ala Leu Gln Ala Ile Ser Glu Gly
            515                 520                 525                 530
Val Gly Thr Ser Leu Leu Ser Thr Leu Ser Ser Pro Gly Pro Lys Leu Asp Asn Ser
        535                 540                 545                 550
Pro Asn Met Asn Ile Thr Gln Pro Ser Lys Val Ser Asn Gln Asp Ser Lys Ser Pro
            555                 560                 565                 570
Leu Gly Phe Tyr Cys Asp Gln Asn Pro Val Glu Ser Ser Met Cys Gln Ser Asn Ser
                575                 580                 585
Arg Asp His Leu Ser Asp Lys Glu Ser Lys Glu Ser Ser Val Glu Gly Ala Glu Asn
590                 595                 600                 605
Gln Arg Gly Pro Leu Glu Ser Lys Gly His Lys Lys Leu Leu Gln Leu Leu Thr Cys
        610                 615                 620                 625
Ser Ser Asp Asp Arg Gly His Ser Ser Leu Thr Asn Ser Pro Leu Asp Ser Ser Cys
        630                 635                 640                 645
Lys Glu Ser Ser Val Ser Val Thr Ser Pro Ser Gly Val Ser Ser Ser Thr Ser Gly
            650                 655                 660                 665
Gly Val Ser Ser Thr Ser Asn Met His Gly Ser Leu Leu Gln Glu Lys His Arg Ile
                670                 675                 680
Leu His Lys Leu Leu Gln Asn Gly Asn Ser Pro Ala Glu Val Ala Lys Ile Thr Ala
685                 690                 695                 700
Glu Ala Thr Gly Lys Asp Thr Ser Ser Ile Thr Ser Cys Gly Asp Gly Asn Val Val
        705                 710                 715                 720
Lys Gln Glu Gln Leu Ser Pro Lys Lys Lys Glu Asn Asn Ala Leu Leu Arg Tyr Leu
        725                 730                 735                 740
Leu Asp Arg Asp Asp Pro Ser Asp Ala Leu Ser Lys Glu Leu Gln Pro Gln Val Glu
            745                 750                 755                 760
Gly Val Asp Asn Lys Met Ser Gln Cys Thr Ser Ser Thr Ile Pro Ser Ser Ser Gln
                765                 770                 775
Glu Lys Asp Pro Lys Ile Lys Thr Glu Thr Ser Glu Glu Gly Ser Gly Asp Leu Asp
780                 785                 790                 795
Asn Leu Asp Ala Ile Leu Gly Asp Leu Thr Ser Ser Asp Phe Tyr Asn Asn Ser Ile
        800                 805                 810                 815
Ser Ser Asn Gly Ser His Leu Gly Thr Lys Gln Gln Val Phe Gln Gly Thr Asn Ser
        820                 825                 830                 835
Leu Gly Leu Lys Ser Ser Gln Ser Val Gln Ser Ile Arg Pro Pro Tyr Asn Arg Ala
            840                 845                 850                 855
Val Ser Leu Asp Ser Pro Val Ser Val Gly Ser Ser Pro Pro Val Lys Asn Ile Ser
                860                 865                 870
Ala Phe Pro Met Leu Pro Lys Gln Pro Met Leu Gly Gly Asn Pro Arg Met Met Asp
875                 880                 885                 890
Ser Gln Glu Asn Tyr Gly Ser Ser Met Gly Gly Pro Asn Arg Asn Val Thr Val Thr
        895                 900                 905                 910
Gln Thr Pro Ser Ser Gly Asp Trp Gly Leu Pro Asn Ser Lys Ala Gly Arg Met Glu
        915                 920                 925                 930
Pro Met Asn Ser Asn Ser Met Gly Arg Pro Gly Gly Asp Tyr Asn Thr Ser Leu Pro
            935                 940                 945                 950
Arg Pro Ala Leu Gly Gly Ser Ile Pro Thr Leu Pro Leu Arg Ser Asn Ser Ile Pro
                955                 960                 965
Gly Ala Arg Pro Val Leu Gln Gln Gln Gln Gln Met Leu Gln Met Arg Pro Gly Glu
970                 975                 980                 985
Ile Pro Met Gly Met Gly Ala Asn Pro Tyr Gly Gln Ala Ala Ala Ser Asn Gln Leu
        990                 995                 1000                1005
Gly Ser Trp Pro Asp Gly Met Leu Ser Met Glu Gln Val Ser His Gly Thr Gln Asn
        1010                1015                1020                1025
Arg Pro Leu Leu Arg Asn Ser Leu Asp Asp Leu Val Gly Pro Pro Ser Asn Leu Glu
            1030                1035                1040
1045
Gly Gln Ser Asp Glu Arg Ala Leu Leu Asp Gln Leu His Thr Leu Leu Ser Asn Thr
                1050                1055                1060
Asp Ala Thr Gly Leu Glu Glu Ile Asp Arg Ala Leu Gly Ile Pro Glu Leu Val Asn
1065                1070                1075                1080
Gln Gly Gln Ala Leu Glu Pro Lys Gln Asp Ala Phe Gln Gly Gln Glu Ala Ala Val
        1085                1090                1095                1100
Met Met Asp Gln Lys Ala Gly Leu Tyr Gly Gln Thr Tyr Pro Ala Gln Gly Pro Pro
        1105                1110                1115                1120
```

```
Met Gln Gly Gly Phe His Leu Gln Gly Gln Ser Pro Ser Phe Asn Ser Met Met Asn
            1125                1130                1135
1140
Gln Met Asn Gln Gln Gly Asn Phe Pro Leu Gln Gly Met His Pro Arg Ala Asn Ile
                1145                1150                1155
Met Arg Pro Arg Thr Asn Thr Pro Lys Gln Leu Arg Met Gln Leu Gln Gln Arg Leu
1160                1165                1170                1175
Gln Gly Gln Gln Phe Leu Asn Gln Ser Arg Gln Ala Leu Glu Leu Lys Met Glu Asn
    1180                1185                1190                1195
Pro Thr Ala Gly Gly Ala Ala Val Met Arg Pro Met Met Gln Pro Gln Gln Gly Phe
        1200                1205                1210                1215
Leu Asn Ala Gln Met Val Ala Gln Arg Ser Arg Glu Leu Leu Ser His His Phe Arg
            1220                1225                1230
1235
Gln Gln Arg Val Ala Met Met Met Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln
                1240                1245                1250
Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Thr
1255                1260                1265                1270
Gln Ala Phe Ser Pro Pro Pro Asn Val Thr Ala Ser Pro Ser Met Asp Gly Leu Leu
    1275                1280                1285                1290
Ala Gly Pro Thr Met Pro Gln Ala Pro Pro Gln Gln Phe Pro Tyr Gln Pro Asn Tyr
        1295                1300                1305                1310
Gly Met Gly Gln Gln Pro Asp Pro Ala Phe Gly Arg Val Ser Ser Pro Pro Asn Ala
            1315                1320                1325                1330
Met Met Ser Ser Arg Met Gly Pro Ser Gln Asn Pro Met Met Gln His Pro Gln Ala
                1335                1340                1345
Ala Ser Ile Tyr Gln Ser Ser Glu Met Lys Gly Trp Pro Ser Gly Asn Leu Ala Arg
1350                1355                1360                1365
Asn Ser Ser Phe Ser Gln Gln Gln Phe Ala His Gln Gly Asn Pro Ala Val Tyr Ser
    1370                1375                1380                1385
Met Val His Met Asn Gly Ser Ser Gly His Met Gly Gln Met Asn Met Asn Pro Met
        1390                1395                1400                1405
Pro Met Ser Gly Met Pro Met Gly Pro Asp Gln Lys Tyr Cys ***
            1410                1415                1420
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22

(B) TYPE: nucleotides

(C) STRANDEDNESS: Single

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

5'-TCATCACTTCCGACAACAGAGG-3'

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20

(B) TYPE: nucleotides

(C) STRANDEDNESS: Single

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

5'-CCAGAAACGTCACTATCAAG-3'

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19

(B) TYPE: nucleotides

(C) STRANDEDNESS: Single

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

5'-TTACTGGAACCCCCATACC-3'

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 950

(B) TYPE: amino acid

(C) STRANDEDNESS: Single

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Cys Ile Gln Arg Phe Phe Ser Leu Asn Asp Gly Gln Ser Trp Ser Gln Lys Arg His
1               5                  10                  15
Tyr Gln Glu Ala Tyr Leu Asn Gly His Ala Glu Thr Pro Val Tyr Arg Phe Ser Leu
            20                  25                  30                  35
Ala Asp Gly Thr Ile Val Thr Ala Gln Thr Lys Ser Lys Leu Phe Arg Asn Pro Val
                40                  45                  50                  55
Thr Asn Asp Arg His Gly Phe Val Ser Thr His Phe Leu Gln Arg Glu Gln Asn Gly
            60                  65                  70                  75
Tyr Arg Pro Asn Pro Asn Pro Val Gly Gln Gly Ile Arg Pro Pro Met Ala Gly Cys
                80                  85                  90                  95
Asn Ser Ser Val Gly Gly Met Ser Met Ser Pro Asn Gln Gly Leu Gln Met Pro Ser
            100                 105                 110
Ser Arg Ala Tyr Gly Leu Ala Asp Pro Ser Thr Thr Gly Gln Met Ser Gly Ala Arg
115                 120                 125                 130
Tyr Gly Gly Ser Ser Asn Ile Ala Ser Leu Thr Pro Gly Pro Gly Met Gln Ser Pro
            135                 140                 145                 150
Ser Ser Tyr Gln Asn Asn Asn Tyr Gly Leu Asn Met Ser Ser Pro Pro His Gly Ser
        155                 160                 165                 170
Pro Gly Leu Ala Pro Asn Gln Gln Asn Ile Met Ile Ser Pro Arg Asn Arg Gly Ser
                175                 180                 185                 190
Pro Lys Ile Ala Ser His Gln Phe Ser Pro Val Ala Gly Val His Ser Pro Met Ala
            195                 200                 205
Ser Ser Gly Asn Thr Gly Asn His Ser Phe Ser Ser Ser Ser Leu Ser Ala Leu Gln
210                 215                 220                 225
Ala Ile Ser Glu Gly Val Gly Thr Ser Leu Leu Ser Thr Leu Ser Ser Pro Gly Pro
            230                 235                 240                 245
Lys Leu Asp Asn Ser Pro Asn Met Asn Ile Thr Gln Pro Ser Lys Val Ser Asn Gln
        250                 255                 260                 265
Asp Ser Lys Ser Pro Leu Gly Phe Tyr Cys Asp Gln Asn Pro Val Glu Ser Ser Met
                270                 275                 280                 285
Cys Gln Ser Asn Ser Arg Asp His Leu Ser Asp Lys Glu Ser Lys Glu Ser Ser Val
            290                 295                 300
Glu Gly Ala Glu Asn Gln Arg Gly Pro Leu Glu Ser Lys Gly His Lys Lys Leu Leu
305                 310                 315                 320
Gln Leu Leu Thr Cys Ser Ser Asp Asp Arg Gly His Ser Ser Leu Thr Asn Ser Pro
            325                 330                 335                 340
Leu Asp Ser Ser Cys Lys Glu Ser Ser Val Ser Val Thr Ser Pro Ser Gly Val Ser
        345                 350                 355                 360
Ser Ser Thr Ser Gly Gly Val Ser Ser Thr Ser Asn Met His Gly Ser Leu Leu Gln
                365                 370                 375                 380
Glu Lys His Arg Ile Leu His Lys Leu Leu Gln Asn Gly Asn Ser Pro Ala Glu Val
            385                 390                 395
Ala Lys Ile Thr Ala Glu Ala Thr Gly Lys Asp Thr Ser Ser Ile Thr Ser Cys Gly
400                 405                 410                 415
Asp Gly Asn Val Val Lys Gln Glu Gln Leu Ser Pro Lys Lys Lys Glu Asn Asn Ala
            420                 425                 430                 435
Leu Leu Arg Tyr Leu Leu Asp Arg Asp Asp Pro Ser Asp Ala Leu Ser Lys Glu Leu
        440                 445                 450                 455
Gln Pro Gln Val Glu Gly Val Asp Asn Lys Met Ser Gln Cys Thr Ser Ser Thr Ile
                460                 465                 470                 475
Pro Ser Ser Ser Gln Glu Lys Asp Pro Lys Ile Lys Thr Glu Thr Ser Glu Glu Gly
            480                 485                 490
Ser Gly Asp Leu Asp Asn Leu Asp Ala Ile Leu Gly Asp Leu Thr Ser Ser Asp Phe
495                 500                 505                 510
Tyr Asn Asn Ser Ile Ser Ser Asn Gly Ser His Leu Gly Thr Lys Gln Gln Val Phe
            515                 520                 525                 530

Gln Gly Thr Asn Ser Leu Gly Leu Lys Ser Ser Gln Ser Val Gln Ser Ile Arg Pro
        535                 540                 545                 550
Pro Tyr Asn Arg Ala Val Ser Leu Asp Ser Pro Val Ser Val Gly Ser Ser Pro Pro
                555                 560                 565                 570
```

```
Val Lys Asn Ile Ser Ala Phe Pro Met Leu Pro Lys Gln Pro Met Leu Gly Gly Asn
                575                 580                 585
Pro Arg Met Met Asp Ser Gln Glu Asn Tyr Gly Ser Ser Met Gly Gly Pro Asn Arg
590                 595                 600                 605
Asn Val Thr Val Thr Gln Thr Pro Ser Ser Gly Asp Trp Gly Leu Pro Asn Ser Lys
    610                 615                 620                 625
Ala Gly Arg Met Glu Pro Met Asn Ser Asn Ser Met Gly Arg Pro Gly Gly Asp Tyr
        630                 635                 640                 645
Asn Thr Ser Leu Pro Arg Pro Ala Leu Gly Gly Ser Ile Pro Thr Leu Pro Leu Arg
            650                 655                 660                 665
Ser Asn Ser Ile Pro Gly Ala Arg Pro Val Leu Gln Gln Gln Gln Gln Met Leu Gln
                670                 675                 680
Met Arg Pro Gly Glu Ile Pro Met Gly Met Gly Ala Asn Pro Tyr Gly Gln Ala Ala
685                 690                 695                 700
Ala Ser Asn Gln Leu Gly Ser Trp Pro Asp Gly Met Leu Ser Met Glu Gln Val Ser
    705                 710                 715                 720
His Gly Thr Gln Asn Arg Pro Leu Leu Arg Asn Ser Leu Asp Asp Leu Val Gly Pro
        725                 730                 735                 740
Pro Ser Asn Leu Glu Gly Gln Ser Asp Glu Arg Ala Leu Leu Asp Gln Leu His Thr
            745                 750                 755                 760
Leu Leu Ser Asn Thr Asp Ala Thr Gly Leu Glu Glu Ile Asp Arg Ala Leu Gly Ile
                765                 770                 775
Pro Glu Leu Val Asn Gln Gly Gln Ala Leu Glu Pro Lys Gln Asp Ala Phe Gln Gly
780                 785                 790                 795
Gln Glu Ala Ala Val Met Met Asp Gln Lys Ala Gly Leu Tyr Gly Gln Thr Tyr Pro
    800                 805                 810                 815
Ala Gln Gly Pro Pro Met Gln Gly Gly Phe His Leu Gln Gly Gln Ser Pro Ser Phe
        820                 825                 830                 835
Asn Ser Met Met Asn Gln Met Asn Gln Gln Gly Asn Phe Pro Leu Gln Gly Met His
            840                 845                 850                 855
Pro Arg Ala Asn Ile Met Arg Pro Arg Thr Asn Thr Pro Lys Gln Leu Arg Met Gln
                860                 865                 870
Leu Gln Gln Arg Leu Gln Gly Gln Gln Phe Leu Asn Gln Ser Arg Gln Ala Leu Glu
875                 880                 885                 890
Leu Lys Met Glu Asn Pro Thr Ala Gly Gly Ala Ala Val Met Arg Pro Met Met Gln
    895                 900                 905                 910
Pro Gln Gln Gly Phe Leu Asn Ala Gln Met Val Ala Gln Arg Ser Arg Glu Leu Leu
        915                 920                 925                 930
Ser His His Phe Arg Gln Gln Arg Val Ala Met Met Met Gln Gln Gln Gln Gln Gln
            935                 940                 945                 950
Gln
```

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4621 nucleotides; 1539 amino acid residues

(B) TYPE: mouse DNA and amino acid

(C) STRANDEDNESS: Single

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
G GCG GCG AAC GGA TCA AAA GAA TTT GCT GAA CAG TGG ACT CCG AGA TCG GTA AAA
1           5                  10                  15
CGA ACT CTT CCC TGC CCT TCC TGA ACA GCT GTC AGT TGC TGA TCT GTG ATC AGG
20                  25                  30                  35
ATG AGT GGA CTA GGC GAA AGC TCT TTG GAT CCG CTG GCC GCT GAG TCT CGG AAA
Met Ser Gly Leu Gly Glu Ser Ser Leu Asp Pro Leu Ala Ala Glu Ser Arg Lys
        40                  45                  50                  55
CGC AAA CTG CCC TGT GAT GCC CCA GGA CAG GGG CTT GTC TAC AGT GGT GAG AAG
Arg Lys Leu Pro Cys Asp Ala Pro Gly Gln Gly Leu Val Tyr Ser Gly Glu Lys
            60                  65                  70
TGG CGA CGG GAG CAG GAG AGC AAG TAC ATA GAG GAG CTG GCA GAG CTC ATC TCT
Trp Arg Arg Glu Gln Glu Ser Lys Tyr Ile Glu Glu Leu Ala Glu Leu Ile Ser
    75                  80                  85                  90
GCA AAT CTC AGC GAC ATC GAC AAC TTC AAT GTC AAG CCA GAT AAA TGT GCC ATC
Ala Asn Leu Ser Asp Ile Asp Asn Phe Asn Val Lys Pro Asp Lys Cys Ala Ile
                95                  100                 105
CTA AAG GAG ACA GTG AGA CAG ATA CGG CAA ATA AAA GAA CAA GGA AAA ACT ATT
Leu Lys Glu Thr Val Arg Gln Ile Arg Gln Ile Lys Glu Gln Gly Lys Thr Ile
110                 115                 120                 125
```

```
TCC AGT GAT GAT GAT GTT CAA AAA GCT GAT GTG TCT TCT ACA GGG CAG GGA GTC
Ser Ser Asp Asp Asp Val Gln Lys Ala Asp Val Ser Ser Thr Gly Gln Gly Val
        130                 135                 140                 145
ATT GAT AAA GAC TCT TTA GGA CCG CTT TTA CTA CAG GCA CTG GAT GGT TTC CTG
Ile Asp Lys Asp Ser Leu Gly Pro Leu Leu Leu Gln Ala Leu Asp Gly Phe Leu
                150                 155                 160
TTT GTG GTG AAT CGA GAT GGA AAC ATT GTA TTC GTG TCA GAA AAT GTC ACA CAG
Phe Val Val Asn Arg Asp Gly Asn Ile Val Phe Val Ser Glu Asn Val Thr Gln
    165                 170                 175                 180
TAT CTG CAG TAC AAG CAG GAG GAC CTG GTT AAC ACA AGT GTC TAC AGC ATC TTA
Tyr Leu Gln Tyr Lys Gln Glu Asp Leu Val Asn Thr Ser Val Tyr Ser Ile Leu
                185                 190                 195
CAT GAG CAA GAC CGG AAG GAT TTT CTT AAA CAC TTA CCA AAA TCC ACA GTT AAT
His Glu Gln Asp Arg Lys Asp Phe Leu Lys His Leu Pro Lys Ser Thr Val Asn
200                 205                 210                 215
GGA GTT TCT TGG ACT AAT GAG AAC CAG AGA CAA AAA AGC CAT ACA TTT AAT TGT
Gly Val Ser Trp Thr Asn Glu Asn Gln Arg Gln Lys Ser His Thr Phe Asn Cys
        220                 225                 230                 235
CGT ATG TTG ATG AAA ACA CAC GAC ATT TTG GAA GAC GTG AAT GCC AGT CCC GAA
Arg Met Leu Met Lys Thr His Asp Ile Leu Glu Asp Val Asn Ala Ser Pro Glu
                240                 245                 250
ACA CGC CAG AGA TAT GAA ACA ATG CAG TGC TTT GCC CTG TCT CAG CCT CGC GCT
Thr Arg Gln Arg Tyr Glu Thr Met Gln Cys Phe Ala Leu Ser Gln Pro Arg Ala
    255                 260                 265                 270
ATG CTG GAA GAA GGA GAA GAC TTG CAG TGC TGT ATG ATC TGC GTG GCT CGC CGC
Met Leu Glu Glu Gly Glu Asp Leu Gln Cys Cys Met Ile Cys Val Ala Arg Arg
            275                 280                 285
GTG ACT GCG CCA TTC CCA TCC AGT CCT GAG AGC TTT ATT ACC AGA CAT GAC CTT
Val Thr Ala Pro Phe Pro Ser Ser Pro Glu Ser Phe Ile Thr Arg His Asp Leu
290                 295                 300                 305
TCC GGA AAG GTT GTC AAT ATA GAT ACA AAC TCA CTT AGA TCT TCC ATG AGG CCT
Ser Gly Lys Val Val Asn Ile Asp Thr Asn Ser Leu Arg Ser Ser Met Arg Pro
        310                 315                 320                 325
GGC TTT GAA GAC ATA ATC CGA AGA TGT ATC CAG AGG TTC TTC AGT CTG AAT GAT
Gly Phe Glu Asp Ile Ile Arg Arg Cys Ile Gln Arg Phe Phe Ser Leu Asn Asp
                330                 335                 340
GGG CAG TCA TGG TCC CAG AAG CGT CAC TAT CAA GAA GCT TAT GTT CAT GGC CAC
Gly Gln Ser Trp Ser Gln Lys Arg His Tyr Gln Glu Ala Tyr Val His Gly His
    345                 350                 355                 360
GCA GAG ACC CCC GTG TAT CGT TTC TCC TTG GCT GAT GGA ACT ATT GTG AGT GCG
Ala Glu Thr Pro Val Tyr Arg Phe Ser Leu Ala Asp Gly Thr Ile Val Ser Ala
            365                 370                 375
CAG ACA AAA AGC AAA CTC TTC CGC AAT CCT GTA ACG AAT GAT CGT CAC GGC TTC
Gln Thr Lys Ser Lys Leu Phe Arg Asn Pro Val Thr Asn Asp Arg His Gly Phe
380                 385                 390                 395
ATC TCG ACC CAC TTT CTT CAG AGA GAA CAG AAT GGA TAC AGA CCA AAC CCA AAT
Ile Ser Thr His Phe Leu Gln Arg Glu Gln Asn Gly Tyr Arg Pro Asn Pro Asn
        400                 405                 410                 415

CCC GCA GGA CAA GGC ATC CGA CCT CCT GCA GCA GGG TGT GGC GTG AGC ATG TCT
Pro Ala Gly Gln Gly Ile Arg Pro Pro Ala Ala Gly Cys Gly Val Ser Met Ser
                420                 425                 430
CCA AAT CAG AAT GTA CAG ATG ATG GGC AGC CGG ACC TAT GGC GTG CCA GAC CCC
Pro Asn Gln Asn Val Gln Met Met Gly Ser Arg Thr Tyr Gly Val Pro Asp Pro
    435                 440                 445                 450
AGC AAC ACA GGG CAG ATG GGT GGA GCT AGG TAC GGG GCT TCT AGT AGC GTA GCC
Ser Asn Thr Gly Gln Met Gly Gly Ala Arg Tyr Gly Ala Ser Ser Ser Val Ala
            455                 460                 465
TCA CTG ACG CCA GGA CAA AGC CTA CAG TCG CCA TCT TCC TAT CAG AAC AGC AGC
Ser Leu Thr Pro Gly Gln Ser Leu Gln Ser Pro Ser Ser Tyr Gln Asn Ser Ser
470                 475                 480                 485
TAT GGG CTC AGC ATG AGC AGT CCC CCC CAC GGC AGT CCT GGT CTT GGT CCC AAC
Tyr Gly Leu Ser Met Ser Ser Pro Pro His Gly Ser Pro Gly Leu Gly Pro Asn
        490                 495                 500                 505
CAG CAG AAC ATC ATG ATT TCC CCT CGG AAT CGT GGC AGC CCA AAG ATG GCC TCC
Gln Gln Asn Ile Met Ile Ser Pro Arg Asn Arg Gly Ser Pro Lys Met Ala Ser
                510                 515                 520
CAC CAG TTC TCT CCT GCT GCA GGT GCA CAC TCA CCC ATG GGA CCT TCT GGC AAC
His Gln Phe Ser Pro Ala Ala Gly Ala His Ser Pro Met Gly Pro Ser Gly Asn
    525                 530                 535                 540
```

```
ACA GGG AGC CAC AGC TTT TCT AGC AGC TCC CTC AGT GCC TTG CAA GCC ATC AGT
Thr Gly Ser His Ser Phe Ser Ser Ser Ser Leu Ser Ala Leu Gln Ala Ile Ser
            545                 550                 555
GAA GGC GTG GGG ACC TCT CTT TTA TCT ACT CTG TCC TCA CCA GGC CCC AAA CTG
Glu Gly Val Gly Thr Ser Leu Leu Ser Thr Leu Ser Ser Pro Gly Pro Lys Leu
560                 565                 570                 575
GAT AAT TCT CCC AAT ATG AAT ATA AGC CAG CCA AGT AAA GTG AGT GGT CAG GAC
Asp Asn Ser Pro Asn Met Asn Ile Ser Gln Pro Ser Lys Val Ser Gly Gln Asp
            580                 585                 590                 595
TCT AAG AGC CCC CTA GGC TTA TAC TGT GAA CAG AAT CCA GTG GAG AGT TCA GTG
Ser Lys Ser Pro Leu Gly Leu Tyr Cys Glu Gln Asn Pro Val Glu Ser Ser Val
                600                 605                 610
TGT CAG TCA AAC AGC AGA GAT CAC CCA AGT GAA AAA GAA AGC AAG GAG AGC AGT
Cys Gln Ser Asn Ser Arg Asp His Pro Ser Glu Lys Glu Ser Lys Glu Ser Ser
    615                 620                 625                 630
GGG GAG GTG TCA GAG ACG CCC AGG GGA CCT CTG GAA AGC AAA GGC CAC AAG AAA
Gly Glu Val Ser Glu Thr Pro Arg Gly Pro Leu Glu Ser Lys Gly His Lys Lys
            635                 640                 645
CTG CTG CAG TTA CTC ACG TGC TCC TCC GAC GAC CGA GGC CAT TCC TCC TTG ACC
Leu Leu Gln Leu Leu Thr Cys Ser Ser Asp Asp Arg Gly His Ser Ser Leu Thr
650                 655                 660                 665
AAC TCT CCC CTG GAT CCA AAC TGC AAA GAC TCT TCC GTT AGT GTC ACC AGC CCC
Asn Ser Pro Leu Asp Pro Asn Cys Lys Asp Ser Ser Val Ser Val Thr Ser Pro
            670                 675                 680                 685
TCT GGA GTG TCC TCC TCA ACA TCA GGG ACA GTG TCT TCC ACC TCC AAT GTG CAT
Ser Gly Val Ser Ser Ser Thr Ser Gly Thr Val Ser Ser Thr Ser Asn Val His
                690                 695                 700
GGG TCT CTG TTG CAA GAG AAA CAC CGG ATT TTG CAC AAG TTG CTG CAG AAT GGC
Gly Ser Leu Leu Gln Glu Lys His Arg Ile Leu His Lys Leu Leu Gln Asn Gly
705                 710                 715                 720
AAC TCC CCA GCG GAG GTC GCC AAG ATC ACT GCA GAG GCC ACT GGG AAG GAC ACG
Asn Ser Pro Ala Glu Val Ala Lys Ile Thr Ala Glu Ala Thr Gly Lys Asp Thr
            725                 730                 735                 740
AGC AGC ACT GCT TCC TGT GGA GAG GGG ACA ACC AGG CAG GAG CAG CTG AGT CCT
Ser Ser Thr Ala Ser Cys Gly Glu Gly Thr Thr Arg Gln Glu Gln Leu Ser Pro
                745                 750                 755
AAG AAG AAG GAG AAT AAT GCT CTG CTT AGA TAC CTG CTG GAC AGG GAT GAC CCC
Lys Lys Lys Glu Asn Asn Ala Leu Leu Arg Tyr Leu Leu Asp Arg Asp Asp Pro
    760                 765                 770                 775
AGT GAT GTG CTT GCC AAA GAG CTG CAG CCC CAG GCC GAC AGT GGG GAC AGT AAA
Ser Asp Val Leu Ala Lys Glu Leu Gln Pro Gln Ala Asp Ser Gly Asp Ser Lys
            780                 785                 790

CTG AGT CAG TGC AGC TGC TCC ACC AAT CCC AGC TCT GGC CAA GAG AAA GAC CCC
Leu Ser Gln Cys Ser Cys Ser Thr Asn Pro Ser Ser Gly Gln Glu Lys Asp Pro
795                 800                 805                 810
AAA ATT AAG ACC GAG ACG AAC GAG GAG GTA TCG GGA GAC CTG GAT AAT CTA GAT
Lys Ile Lys Thr Glu Thr Asn Glu Glu Val Ser Gly Asp Leu Asp Asn Leu Asp
            815                 820                 825                 830
GCC ATT CTT GGA GAT TTG ACC AGT TCT GAC TTC TAC AAC AAT CCT ACA AAT GGC
Ala Ile Leu Gly Asp Leu Thr Ser Ser Asp Phe Tyr Asn Asn Pro Thr Asn Gly
                835                 840                 845
GGT CAC CCA GGG GCC AAA CAG CAG ATG TTT GCA GGA CCG AGT TCT CTG GGT TTG
Gly His Pro Gly Ala Lys Gln Gln Met Phe Ala Gly Pro Ser Ser Leu Gly Leu
    850                 855                 860                 865
CGA AGT CCA CAG CCT GTG CAG TCT GTT CGT CCT CCA TAT AAC CGA GCG GTG TCT
Arg Ser Pro Gln Pro Val Gln Ser Val Arg Pro Pro Tyr Asn Arg Ala Val Ser
            870                 875                 880
CTG GAT AGC CCT GTG TCT GTT GGC TCA GGT CCG CCA GTG AAG AAT GTC AGT GCT
Leu Asp Ser Pro Val Ser Val Gly Ser Gly Pro Pro Val Lys Asn Val Ser Ala
885                 890                 895                 900
TTC CCT GGG TTA CCA AAA CAG CCC ATA CTG GCT GGG AAT CCA AGA ATG ATG GAT
Phe Pro Gly Leu Pro Lys Gln Pro Ile Leu Ala Gly Asn Pro Arg Met Met Asp
            905                 910                 915                 920
AGT CAG GAG AAT TAC GGT GCC AAC ATG GGC CCA AAC AGA AAT GTT CCT GTG AAT
Ser Gln Glu Asn Tyr Gly Ala Asn Met Gly Pro Asn Arg Asn Val Pro Val Asn
                925                 930                 935
CCG ACT TCC TCC CCC GGA GAC TGG GGC TTA GCT AAC TCA AGG GCC AGC AGA ATG
Pro Thr Ser Ser Pro Gly Asp Trp Gly Leu Ala Asn Ser Arg Ala Ser Arg Met
940                 945                 950                 955
GAG CCT CTG GCA TCA AGT CCC CTG GGA AGA ACT GGA GCC GAT TAC AGT GCC ACT
```

```
Glu Pro Leu Ala Ser Ser Pro Leu Gly Arg Thr Gly Ala Asp Tyr Ser Ala Thr
        960                 965                 970                 975
TTA CCC AGA CCT GCC ATG GGG GGC TCT GTG CCT ACC TTG CCA CTT CGT TCT AAT
Leu Pro Arg Pro Ala Met Gly Gly Ser Val Pro Thr Leu Pro Leu Arg Ser Asn
                980                 985                 990
CGA CTG CCA GGT GCA AGA CCA TCG TTG CAG CAA CAG CAG CAG CAA CAG CAG CAA
Arg Leu Pro Gly Ala Arg Pro Ser Leu Gln Gln Gln Gln Gln Gln Gln Gln Gln
    995                 1000                1005                1010
CAG CAA CAA CAA CAG CAG CAA CAG CAG CAG CAA CAG CAG CAG CAG CAA CAG CAG
Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln
            1015                1020                1025
CAG ATG CTT CAA ATG AGA ACT GGT GAG ATT CCC ATG GGA ATG GGA GTC AAT CCC
Gln Met Leu Gln Met Arg Thr Gly Glu Ile Pro Met Gly Met Gly Val Asn Pro
1030                1035                1040                1045
TAT AGC CCA GCA GTG CCG TCT AAC CAA CCA GGT TCC TGG CCA GAG GGC ATG CTC
Tyr Ser Pro Ala Val Pro Ser Asn Gln Pro Gly Ser Trp Pro Glu Gly Met Leu
        1050                1055                1060                1065
TCT ATG GAA CAA GGT CCT CAC GGG TCT CAA AAT AGG CCT CTT CTT AGA AAC TCT
Ser Met Glu Gln Gly Pro His Gly Ser Gln Asn Arg Pro Leu Leu Arg Asn Ser
                1070                1075                1080
CTG GAT GAT CTG CTT GGG CCA CCT TCT AAC GCA GAG GGC CAG AGT GAC GAG AGA
Leu Asp Asp Leu Leu Gly Pro Pro Ser Asn Ala Glu Gly Gln Ser Asp Glu Arg
    1085                1090                1095                1100
GCT CTG CTG GAC CAG CTG CAC ACA CTC CTG AGC AAC ACA GAT GCC ACA GGT CTG
Ala Leu Leu Asp Gln Leu His Thr Leu Leu Ser Asn Thr Asp Ala Thr Gly Leu
            1105                1110                1115
GAG GAG ATC GAC AGG GCC TTG GGA ATT CCT GAG CTC GTG AAT CAG GGA CAA GCT
Glu Glu Ile Asp Arg Ala Leu Gly Ile Pro Glu Leu Val Asn Gln Gly Gln Ala
1120                1125                1130                1135
TTG GAG TCC AAA CAG GAT GTT TTC CAA GGC CAA GAA GCA GCA GTA ATG ATG GAT
Leu Glu Ser Lys Gln Asp Val Phe Gln Gly Gln Glu Ala Ala Val Met Met Asp
        1140                1145                1150                1155
CAG AAG GCT GCA CTA TAT GGA CAG ACA TAC CCA GCT CAG GGT CCT CCC CTT CAA
Gln Lys Ala Ala Leu Tyr Gly Gln Thr Tyr Pro Ala Gln Gly Pro Pro Leu Gln
                1160                1165                1170

GGA GGC TTT AAC CTT CAG GGA CAG TCA CCA TCG TTT AAC TCT ATG ATG GGT CAG
Gly Gly Phe Asn Leu Gln Gly Gln Ser Pro Ser Phe Asn Ser Met Met Gly Gln
    1175                1180                1185                1190
ATT AGC CAG CAA GGC AGC TTT CCT CTG CAA GGC ATG CAT CCT AGA GCC GGC CTC
Ile Ser Gln Gln Gly Ser Phe Pro Leu Gln Gly Met His Pro Arg Ala Gly Leu
            1195                1200                1205
GTG AGA CCA AGG ACC AAC ACC CCG AAG CAG CTG AGA ATG CAG CTT CAG CAG AGG
Val Arg Pro Arg Thr Asn Thr Pro Lys Gln Leu Arg Met Gln Leu Gln Gln Arg
1210                1215                1220                1225
CTA CAG GGC CAG CAG TTT TTA AAT CAG AGC CGG CAG GCA CTT GAA ATG AAA ATG
Leu Gln Gly Gln Gln Phe Leu Asn Gln Ser Arg Gln Ala Leu Glu Met Lys Met
        1230                1235                1240                1245
GAG AAC CCT GCT GGC ACT GCT GTG ATG AGG CCC ATG ATG CCC CAG GCT TTC TTT
Glu Asn Pro Ala Gly Thr Ala Val Met Arg Pro Met Met Pro Gln Ala Phe Phe
                1250                1255                1260
AAT GCC CAA ATG GCT GCC CAG CAG AAA CGA GAG CTG ATG AGC CAT CAC CTG CAG
Asn Ala Gln Met Ala Ala Gln Gln Lys Arg Glu Leu Met Ser His His Leu Gln
    1265                1270                1275                1280
CAG CAG AGG ATG GCG ATG ATG ATG TCA CAA CCA CAG CCT CAG GCC TTC AGC CCA
Gln Gln Arg Met Ala Met Met Met Ser Gln Pro Gln Pro Gln Ala Phe Ser Pro
            1285                1290                1295
CCT CCC AAC GTC ACC GCC TCC CCC AGC ATG GAC GGG GTT TTG GCA GGT TCA GCA
Pro Pro Asn Val Thr Ala Ser Pro Ser Met Asp Gly Val Leu Ala Gly Ser Ala
1300                1305                1310                1315
ATG CCG CAA GCC CCT CCA CAA CAG TTT CCA TAT CCA GCA AAT TAC GGA ATG GGA
Met Pro Gln Ala Pro Pro Gln Gln Phe Pro Tyr Pro Ala Asn Tyr Gly Met Gly
        1320                1325                1330                1335
CAA CCA CCA GAG CCA GCC TTT GGT CGA GGC TCG AGT CCT CCC AGT GCA ATG ATG
Gln Pro Pro Glu Pro Ala Phe Gly Arg Gly Ser Ser Pro Pro Ser Ala Met Met
                1340                1345                1350
TCA TCA AGA ATG GGG CCT TCC CAG AAT GCC ATG GTG CAG CAT CCT CAG CCC ACA
Ser Ser Arg Met Gly Pro Ser Gln Asn Ala Met Val Gln His Pro Gln Pro Thr
    1355                1360                1365                1370
CCC ATG TAT CAG CCT TCA GAT ATG AAG GGG TGG CCG TCA GGG AAC CTG GCC AGG
Pro Met Tyr Gln Pro Ser Asp Met Lys Gly Trp Pro Ser Gly Asn Leu Ala Arg
```

```
          1375                1380                1385
AAT GGC TCC TTC CCC CAG CAG CAG TTT GCT CCC CAG GGG AAC CCT GCA GCC TAC
Asn Gly Ser Phe Pro Gln Gln Gln Phe Ala Pro Gln Gly Asn Pro Ala Ala Tyr
     1390                1395                1400                1405
AAC ATG GTG CAT ATG AAC AGC AGC GGT GGG CAC TTG GGA CAG ATG GCC ATG ACC
Asn Met Val His Met Asn Ser Ser Gly Gly His Leu Gly Gln Met Ala Met Thr
               1410                1415                1420
CCC ATG CCC ATG TCT GGC ATG CCC ATG GGC CCC GAT CAG AAA TAC TGC TGA CAT
Pro Met Pro Met Ser Gly Met Pro Met Gly Pro Asp Gln Lys Tyr Cys *** His
     1425                1430                1435                1440
CTC CCT AGT GGG ACT GAC TGT ACA GAT GAC ACT GCA CAG GAT CAT CAG GAC GTG
Leu Pro Ser Gly Thr Asp Cys Thr Asp Asp Thr Ala Gln Asp His Gln Asp Val
               1445                1450                1455
GCG GCG AGT CAT TGT CTA AGC ATC CAG CTT GGA AAC AAG GCC AGC GTG ACC AGC
Ala Ala Ser His Cys Leu Ser Ile Gln Leu Gly Asn Lys Ala Ser Val Thr Ser
1460                1465                1470                1475
AGC GGG GTC TGT GCT GTC ATT TGA GCA GAG CTG GGT CTC GCT GAA GCG CAC TGT
Ser Gly Val Cys Ala Val Ile ***
          1480                1485                1490                1495
CTA CCT GAT GCC CTG CCT CTG TGT GGC AAG GTG TTC TGC CTC ATG AGG ATG TGA
               1500                1505                1510
TTC TGG AGA TGG GGT GTT CGT AAG CAC CGC TCT CTT ACG TCA CTC CCT TCT GCC
     1515                1520                1525                1530
TCG CCA GCC AAA GTC TTC ACG TAG ATC TAG
               1535                1540
```

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22

(B) TYPE: nucleic acid

(C) STRANDEDNESS: Single

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

5'-TCCTTTTCCCAGCAGCAGTTTG-3'

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20

(B) TYPE: nucleic acid

(C) STRANDEDNESS: Single

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

5'ATGCCAGACATGGGCATGGG-3'

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1539

(B) TYPE: amino acid

(C) STRANDEDNESS: Single

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
Met Ser Gly Leu Gly Glu Ser Ser Leu Asp Pro Leu Ala Ala Glu Ser Arg Lys
          40                  45                  50                  55
Arg Lys Leu Pro Cys Asp Ala Pro Gly Gln Gly Leu Val Tyr Ser Gly Glu Lys
                    60                  65                  70
Trp Arg Arg Glu Gln Glu Ser Lys Tyr Ile Glu Glu Leu Ala Glu Leu Ile Ser
     75                  80                  85                  90
Ala Asn Leu Ser Asp Ile Asp Asn Phe Asn Val Lys Pro Asp Lys Cys Ala Ile
               95                  100                 105
Leu Lys Glu Thr Val Arg Gln Ile Arg Gln Ile Lys Glu Gln Gly Lys Thr Ile
110                 115                 120                 125
```

```
Ser Ser Asp Asp Asp Val Gln Lys Ala Asp Val Ser Ser Thr Gly Gln Gly Val
        130                 135                 140                 145
Ile Asp Lys Asp Ser Leu Gly Pro Leu Leu Leu Gln Ala Leu Asp Gly Phe Leu
                    150                 155                 160
Phe Val Val Asn Arg Asp Gly Asn Ile Val Phe Val Ser Glu Asn Val Thr Gln
    165                 170                 175                 180
Tyr Leu Gln Tyr Lys Gln Glu Asp Leu Val Asn Thr Ser Val Tyr Ser Ile Leu
                185                 190                 195
His Glu Gln Asp Arg Lys Asp Phe Leu Lys His Leu Pro Lys Ser Thr Val Asn
200                 205                 210                 215
Gly Val Ser Trp Thr Asn Glu Asn Gln Arg Gln Lys Ser His Thr Phe Asn Cys
        220                 225                 230                 235
Arg Met Leu Met Lys Thr His Asp Ile Leu Glu Asp Val Asn Ala Ser Pro Glu
                240                 245                 250
Thr Arg Gln Arg Tyr Glu Thr Met Gln Cys Phe Ala Leu Ser Gln Pro Arg Ala
    255                 260                 265                 270
Met Leu Glu Glu Gly Glu Asp Leu Gln Cys Cys Met Ile Cys Val Ala Arg Arg
                275                 280                 285
Val Thr Ala Pro Phe Pro Ser Ser Pro Glu Ser Phe Ile Thr Arg His Asp Leu
290                 295                 300                 305

Ser Gly Lys Val Val Asn Ile Asp Thr Asn Ser Leu Arg Ser Ser Met Arg Pro
        310                 315                 320                 325
Gly Phe Glu Asp Ile Ile Arg Arg Cys Ile Gln Arg Phe Phe Ser Leu Asn Asp
                330                 335                 340
Gly Gln Ser Trp Ser Gln Lys Arg His Tyr Gln Glu Ala Tyr Val His Gly His
    345                 350                 355                 360
Ala Glu Thr Pro Val Tyr Arg Phe Ser Leu Ala Asp Gly Thr Ile Val Ser Ala
                365                 370                 375
Gln Thr Lys Ser Lys Leu Phe Arg Asn Pro Val Thr Asn Asp Arg His Gly Phe
380                 385                 390                 395
Ile Ser Thr His Phe Leu Gln Arg Glu Gln Asn Gly Tyr Arg Pro Asn Pro Asn
        400                 405                 410                 415
Pro Ala Gly Gln Gly Ile Arg Pro Pro Ala Ala Gly Cys Gly Val Ser Met Ser
                420                 425                 430
Pro Asn Gln Asn Val Gln Met Met Gly Ser Arg Thr Tyr Gly Val Pro Asp Pro
    435                 440                 445                 450
Ser Asn Thr Gly Gln Met Gly Gly Ala Arg Tyr Gly Ala Ser Ser Ser Val Ala
                455                 460                 465
Ser Leu Thr Pro Gly Gln Ser Leu Gln Ser Pro Ser Ser Tyr Gln Asn Ser Ser
470                 475                 480                 485
Tyr Gly Leu Ser Met Ser Ser Pro Pro His Gly Ser Pro Gly Leu Gly Pro Asn
        490                 495                 500                 505
Gln Gln Asn Ile Met Ile Ser Pro Arg Asn Arg Gly Ser Pro Lys Met Ala Ser
                510                 515                 520
His Gln Phe Ser Pro Ala Ala Gly Ala His Ser Pro Met Gly Pro Ser Gly Asn
    525                 530                 535                 540
Thr Gly Ser His Ser Phe Ser Ser Ser Ser Leu Ser Ala Leu Gln Ala Ile Ser
                545                 550                 555
Glu Gly Val Gly Thr Ser Leu Leu Ser Thr Leu Ser Ser Pro Gly Pro Lys Leu
560                 565                 570                 575
Asp Asn Ser Pro Asn Met Asn Ile Ser Gln Pro Ser Lys Val Ser Gly Gln Asp
        580                 585                 590                 595
Ser Lys Ser Pro Leu Gly Leu Tyr Cys Glu Gln Asn Pro Val Glu Ser Ser Val
                600                 605                 610
Cys Gln Ser Asn Ser Arg Asp His Pro Ser Glu Lys Glu Ser Lys Glu Ser Ser
    615                 620                 625                 630
Gly Glu Val Ser Glu Thr Pro Arg Gly Pro Leu Glu Ser Lys Gly His Lys Lys
                635                 640                 645
Leu Leu Gln Leu Leu Thr Cys Ser Ser Asp Asp Arg Gly His Ser Ser Leu Thr
650                 655                 660                 665
Asn Ser Pro Leu Asp Pro Asn Cys Lys Asp Ser Ser Val Ser Val Thr Ser Pro
        670                 675                 680                 685
Ser Gly Val Ser Ser Ser Thr Ser Gly Thr Val Ser Ser Thr Ser Asn Val His
                690                 695                 700
Gly Ser Leu Leu Gln Glu Lys His Arg Ile Leu His Lys Leu Leu Gln Asn Gly
705                 710                 715                 720
Asn Ser Pro Ala Glu Val Ala Lys Ile Thr Ala Glu Ala Thr Gly Lys Asp Thr
        725                 730                 735                 740
Ser Ser Thr Ala Ser Cys Gly Glu Gly Thr Thr Arg Gln Glu Gln Leu Ser Pro
                745                 750                 755
```

Lys Lys Lys Glu Asn Asn Ala Leu Leu Arg Tyr Leu Leu Asp Arg Asp Asp Pro
760 765 770 775
Ser Asp Val Leu Ala Lys Glu Leu Gln Pro Gln Ala Asp Ser Gly Asp Ser Lys
780 785 790
Leu Ser Gln Cys Ser Cys Ser Thr Asn Pro Ser Ser Gly Gln Glu Lys Asp Pro
795 800 805 810
Lys Ile Lys Thr Glu Thr Asn Glu Glu Val Ser Gly Asp Leu Asp Asn Leu Asp
815 820 825 830
Ala Ile Leu Gly Asp Leu Thr Ser Ser Asp Phe Tyr Asn Asn Pro Thr Asn Gly
835 840 845
Gly His Pro Gly Ala Lys Gln Gln Met Phe Ala Gly Pro Ser Ser Leu Gly Leu
850 855 860 865
Arg Ser Pro Gln Pro Val Gln Ser Val Arg Pro Pro Tyr Asn Arg Ala Val Ser
870 875 880
Leu Asp Ser Pro Val Ser Val Gly Ser Gly Pro Pro Val Lys Asn Val Ser Ala
885 890 895 900
Phe Pro Gly Leu Pro Lys Gln Pro Ile Leu Ala Gly Asn Pro Arg Met Met Asp
905 910 915 920
Ser Gln Glu Asn Tyr Gly Ala Asn Met Gly Pro Asn Arg Asn Val Pro Val Asn
925 930 935
Pro Thr Ser Ser Pro Gly Asp Trp Gly Leu Ala Asn Ser Arg Ala Ser Arg Met
940 945 950 955
Glu Pro Leu Ala Ser Ser Pro Leu Gly Arg Thr Gly Ala Asp Tyr Ser Ala Thr
960 965 970 975
Leu Pro Arg Pro Ala Met Gly Gly Ser Val Pro Thr Leu Pro Leu Arg Ser Asn
980 985 990
Arg Leu Pro Gly Ala Arg Pro Ser Leu Gln Gln Gln Gln Gln Gln Gln Gln Gln
995 1000 1005 1010
Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln
1015 1020 1025
Gln Met Leu Gln Met Arg Thr Gly Glu Ile Pro Met Gly Met Gly Val Asn Pro
1030 1035 1040 1045
Tyr Ser Pro Ala Val Pro Ser Asn Gln Pro Gly Ser Trp Pro Glu Gly Met Leu
1050 1055 1060 1065
Ser Met Glu Gln Gly Pro His Gly Ser Gln Asn Arg Pro Leu Leu Arg Asn Ser
1070 1075 1080
Leu Asp Asp Leu Leu Gly Pro Pro Ser Asn Ala Glu Gly Gln Ser Asp Glu Arg
1085 1090 1095 1100
Ala Leu Leu Asp Gln Leu His Thr Leu Leu Ser Asn Thr Asp Ala Thr Gly Leu
1105 1110 1115
Glu Glu Ile Asp Arg Ala Leu Gly Ile Pro Glu Leu Val Asn Gln Gly Gln Ala
1120 1125 1130 1135
Leu Glu Ser Lys Gln Asp Val Phe Gln Gly Gln Glu Ala Ala Val Met Met Asp
1140 1145 1150 1155
Gln Lys Ala Ala Leu Tyr Gly Gln Thr Tyr Pro Ala Gln Gly Pro Pro Leu Gln
1160 1165 1170
Gly Gly Phe Asn Leu Gln Gly Gln Ser Pro Ser Phe Asn Ser Met Met Gly Gln
1175 1180 1185 1190
Ile Ser Gln Gln Gly Ser Phe Pro Leu Gln Gly Met His Pro Arg Ala Gly Leu
1195 1200 1205
Val Arg Pro Arg Thr Asn Thr Pro Lys Gln Leu Arg Met Gln Leu Gln Gln Arg
1210 1215 1220 1225
Leu Gln Gly Gln Gln Phe Leu Asn Gln Ser Arg Gln Ala Leu Glu Met Lys Met
1230 1235 1240 1245
Glu Asn Pro Ala Gly Thr Ala Val Met Arg Pro Met Met Pro Gln Ala Phe Phe
1250 1255 1260
Asn Ala Gln Met Ala Ala Gln Gln Lys Arg Glu Leu Met Ser His His Leu Gln
1265 1270 1275 1280
Gln Gln Arg Met Ala Met Met Met Ser Gln Pro Gln Pro Gln Ala Phe Ser Pro
1285 1290 1295
Pro Pro Asn Val Thr Ala Ser Pro Ser Met Asp Gly Val Leu Ala Gly Ser Ala
1300 1305 1310 1315
Met Pro Gln Ala Pro Pro Gln Gln Phe Pro Tyr Pro Ala Asn Tyr Gly Met Gly
1320 1325 1330 1335
Gln Pro Pro Glu Pro Ala Phe Gly Arg Gly Ser Ser Pro Pro Ser Ala Met Met
1340 1345 1350
Ser Ser Arg Met Gly Pro Ser Gln Asn Ala Met Val Gln His Pro Gln Pro Thr
1355 1360 1365 1370
Pro Met Tyr Gln Pro Ser Asp Met Lys Gly Trp Pro Ser Gly Asn Leu Ala Arg
1375 1380 1385
Asn Gly Ser Phe Pro Gln Gln Gln Phe Ala Pro Gln Gly Asn Pro Ala Ala Tyr

```
     1390                1395                1400                1405
Asn Met Val His Met Asn Ser Ser Gly Gly His Leu Gly Gln Met Ala Met Thr
                1410                1415                1420
Pro Met Pro Met Ser Gly Met Pro Met Gly Pro Asp Gln Lys Tyr Cys *** His
    1425                1430                1435                1440
Leu Pro Ser Gly Thr Asp Cys Thr Asp Asp Thr Ala Gln Asp His Gln Asp Val
            1445                1450                1455
Ala Ala Ser His Cys Leu Ser Ile Gln Leu Gly Asn Lys Ala Ser Val Thr Ser
1460                1465                1470                1475
Ser Gly Val Cys Ala Val Ile ***
        1480                1485                1490                1495
```

What is claimed is:

1. A substantially pure DNA comprising a sequence encoding an AIB1 polypeptide.

2. The DNA of claim 1, wherein the polypeptide is human AIB1.

3. The DNA of claim 1, wherein the polypeptide comprises the amino acid sequence of SEQ. I.D. NO. 4.

4. The DNA of claim 1, wherein the polypeptide comprises the amino acid sequence of SEQ. I.D. NO. 2.

5. The DNA of claim 1, wherein the polypeptide comprises the amino acid sequence of SEQ. I.D. NO. 3.

6. The DNA of claim 1, wherein the polypeptide comprises the amino acid sequence of SEQ. I.D. NO. 8.

7. A substantially pure DNA comprising a polynucleotide which hybridizes at high stringency to a DNA having the sequence of SEQ. I.D. NO. 1, or the complement thereof.

8. A substantially pure DNA comprising a nucleotide sequence having at least 50% sequence identity to SEQ. I.D. NO. 1, the nucleotide sequence encoding a polypeptide having the biological activity of a AIB1 polypeptide.

9. A substantially pure DNA comprising (a) the sequence of SEQ. I.D. NO. 1 or (b) a degenerate variant thereof.

10. The DNA of claim 1, wherein the DNA is operably linked to regulatory sequences for expression of the polypeptide, the regulatory sequences comprising a promoter.

11. A cell comprising the DNA of claim 1.

12. A substantially pure human AIB1 polypeptide.

13. The polypeptide of claim 12, wherein the polypeptide comprises the amino acid sequence of SEQ. I.D. Nos. 2, 3, 4, or 8.

14. A method of identifying a candidate compound which inhibits estrogen receptor (ER)-dependent transcription comprising contacting the compound with an AIB1 polypeptide and determining whether the compound binds to the polypeptide, wherein binding of the compound to the polypeptide indicates that the compound inhibits ER-dependent transcription.

15. The method of claim 14, wherein the AIB1 polypeptide comprises a Per/Arnt/Sim (PAS) domain.

16. The method of claim 14, wherein the AIB1 polypeptide comprises a basic helix-loop-helix (bHLH) domain.

17. The method of claim 14, wherein the AIB1 polypeptide comprises an ER-interacting domain.

18. A method of identifying a candidate compound which inhibits ER-dependent transcription comprising:

contacting the compound with an AIB1 polypeptide and an ER polypeptide and determining the ability of the compound to interfere with the binding of the ER polypeptide with the AIB1 polypeptide.

19. The method of claim 18, wherein the AIB1 polypeptide comprises a PAS domain.

20. The method of claim 18, wherein the AIB1 polypeptide comprises a bHLH domain.

21. A method of screening a candidate compound which inhibits an interaction of an AIB1 polypeptide with an ER polypeptide in a cell comprising

(a) providing a GAL4 binding site linked to a reporter gene;

(b) providing a GAL4 binding domain linked to either (i) an AIB1 polypeptide or (ii) an ER polypeptide;

(c) providing a GAL4 transactivation domain II linked to the ER polypeptide if the GAL4 binding domain is linked to the AIB1 polypeptide or linked to the AIB1 polypeptide if the GAL4 binding domain is linked to the ER polypeptide;

(d) contacting the cell with the compound; and

(e) monitoring expression of the reporter gene, wherein a decrease in expression in the presence of the compound compared to that in the absence of the compound indicates that the compound inhibits an interaction of an AIB1 polypeptide with the ER polypeptide.

22. A method of detecting an aberrantly proliferating cell in a tissue sample comprising determining the level of AIB1 gene expression in the sample, wherein an increase in the level of expression compared to the level in normal control tissue indicates the presence of an aberrantly proliferating cell.

23. The method of claim 21, wherein the aberrantly proliferating cell is a steroid hormone-responsive cancer cell.

24. The method of claim 23, wherein the steroid hormone-responsive cancer cell is a breast cancer cell.

25. The method of claim 23, wherein the cell is a steroid hormone-responsive cancer cell is an ovarian cancer cell.

26. The method of claim 21, wherein the AIB1 gene expression is measured using an AIB1 gene-specific polynucleotide probe.

27. The method of claim 21, wherein the AIB1 gene expression is measured using an antibody specific for an AIB1 gene product.

28. A method of detecting breast cancer in a tissue sample, comprising determining the number of cellular copies of an AIB1 gene in the tissue sample, wherein an increase in the number of copies compared to the number of copies in a normal control tissue indicates the presence of a breast carcinoma.

29. The method of claim 28, wherein the number of copies in the tissue is greater than 2.

30. The method of claim 29, wherein the number of copies in the tissue is greater than 10.

31. The method of claim 30, wherein the number of copies in the tissue is greater than 20.

32. A method of reducing proliferation of a cancer cell in a mammal comprising administering to the mammal a compound which inhibits expression of AIB1.

33. The method of claim 32, wherein the compound reduces transcription of DNA encoding AIB1 in the cell.

34. The method of claim 32, wherein the compound reduces translation of an AIB1 mRNA into an AIB1 gene product in the cell.

35. The method of claim 34, wherein the translation is reduced by contacting the AIB1 mRNA with an antisense DNA complementary to the AIB1 mRNA.

36. A method of inhibiting ER-dependent transcription in a breast cell of an mammal, comprising administering an effective amount of an AIB1 polypeptide to the mammal.

37. The method of claim 36, wherein the polypeptide comprises a PAS domain.

38. The method of claim 36, wherein the polypeptide comprises a bHLH domain.

39. The method of claim 36, wherein the polypeptide comprises an ER-interacting domain

40. A method of inhibiting ER-dependent transcription in a cancer cell of a mammal, comprising administering an effective amount of a peptide mimetic of an AIB1 polypeptide to the mammal.

41. A monoclonal antibody which binds specifically to AIB1.

42. A method of identifying a tamoxifen-sensitive patient, comprising

(a) contacting a patient-derived tissue sample with tamoxifen; and

(b) determining the level of AIB1 gene expression in the sample, wherein an increase in the level of expression compared to the level in normal control tissue indicates that the patient is tamoxifen-sensitive.

43. The method of claim 42, wherein the AIB1 gene expression is measured using an AIB1 gene-specific polynucleotide probe.

44. The method of claim 42, wherein the AIB1 gene expression is measured using an antibody specific for an AIB1 gene product.

45. A transgenic animal wherein at least one copy of the AIB1 gene has been functionally deleted.

46. A transgenic mouse wherein at least one copy of the pCIP gene has been functionally deleted.

47. The invention of claim 45 wherein at least one copy of the gene has been functionally deleted using a method selected from the group consisting of: anti-sense technology, transposon mutagenesis, homologous recombination with a non-functional gene homolog of AIB1.

48. A transgenic animal genetically engineered to have more than the normal copy number of the AIB1 gene.

49. The invention of claim 48 wherein at least one copy of the AIB1 gene has been introduced into the animal on an extra-chromosomal element.

50. A transgenic animal having at least one AIB1 gene operatively linked to a non-native promoter.

51. The invention of claim 50 wherein the non-native promoter is selected from the group consisting of: a mouse mammary tumor virus promoter, a whey acidic protein promoter and a metallothionein promoter.

52. The invention of claim 50 wherein transcription from the promoter has the characteristic selected from the group consisting of: being inducible, being repressible and being constitutive.

53. A method of reducing proliferation of a cancer cell comprising administering to the mammal a compound which inhibits interaction of AIB1 with a molecule selected from the group consisting of steroid receptors and nuclear co-factors.

54. The method of claim 58 wherein the molecule is selected from the group consisting of: p300 and CBP.

AIB1, A NOVEL STEROID RECEPTOR CO-ACTIVATOR

Abstract of the Disclosure

The invention features a substantially pure DNA which includes a sequence encoding a novel steroid receptor co-activator which is overexpressed in breast cancer cells, diagnostic assays for steroid hormone-responsive cancers, and screening assays to identify compounds which inhibit an interaction of the co-activator with the steroid hormone.

Figure 1

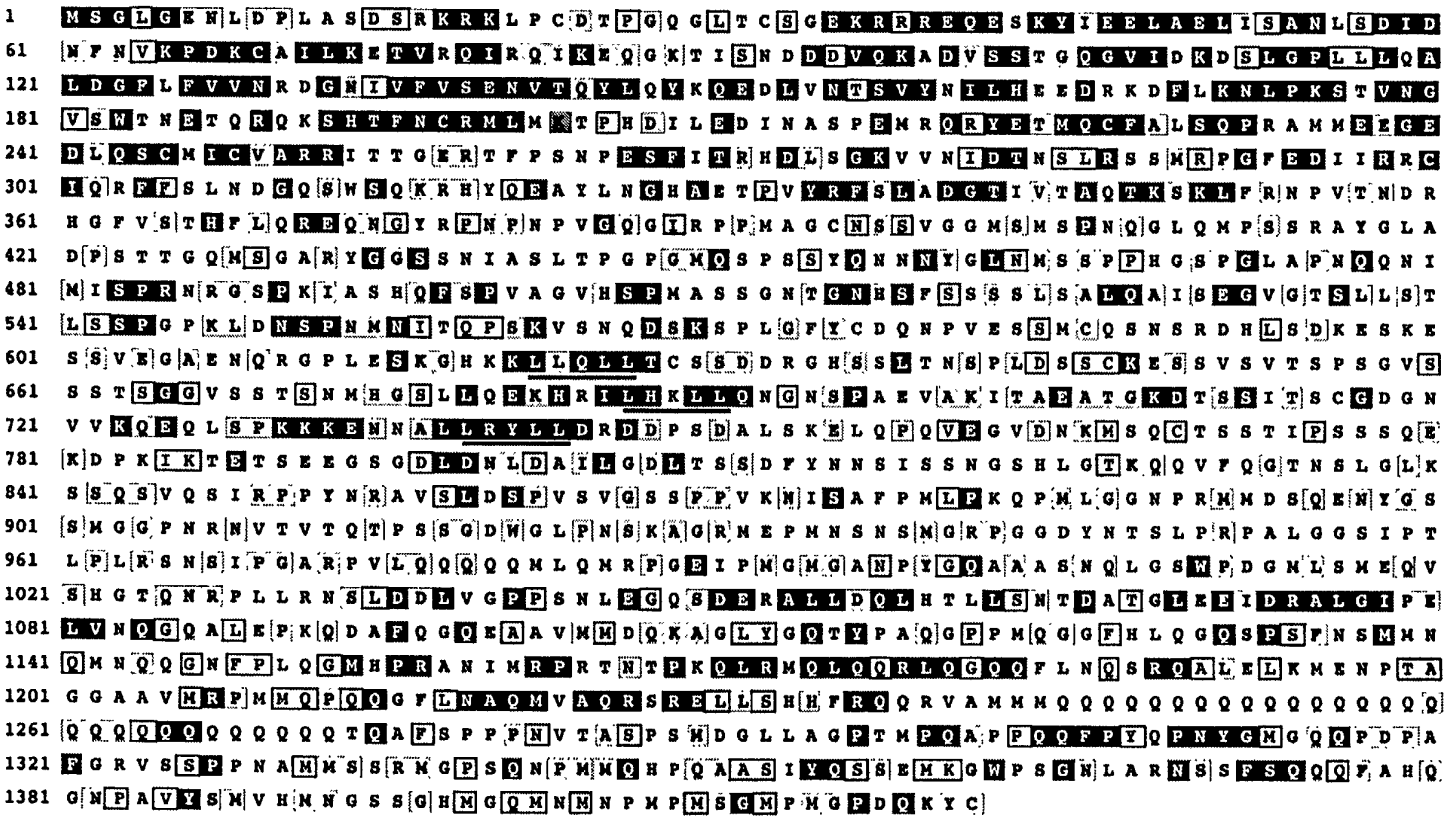

Figure 2

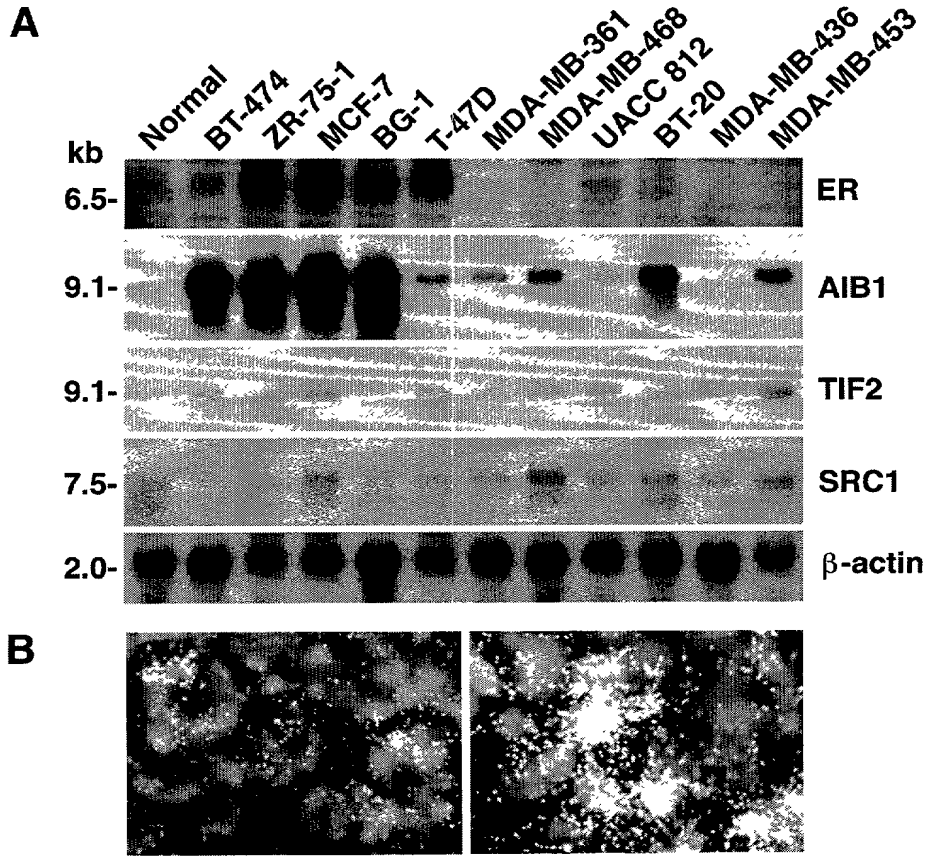
A
Normal
BT-474
ZR-75-1
MCF-7
BG-1
T-47D
MDA-MB-361
MDA-MB-468
UACC 812
BT-20
MDA-MB-436
MDA-MB-453
kb
6.5-
ER
9.1-
AIB1
9.1-
TIF2
7.5-
SRC1
2.0-
β-actin
B

Figure 3

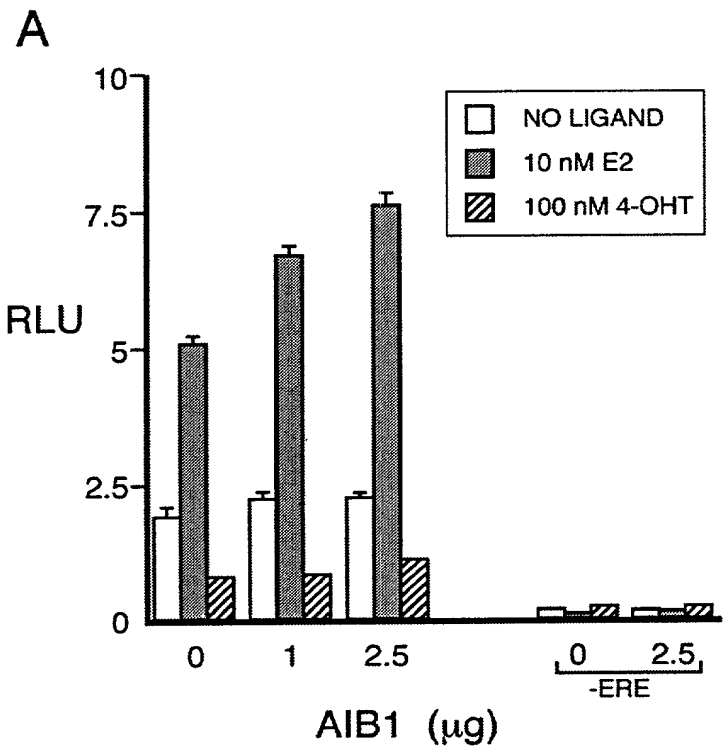
A
10
7.5
5
2.5
0
RLU
NO LIGAND
10 nM E2
100 nM 4-OHT
0
1
2.5
0
2.5
-ERE
AIB1 (μg)

Figure 4

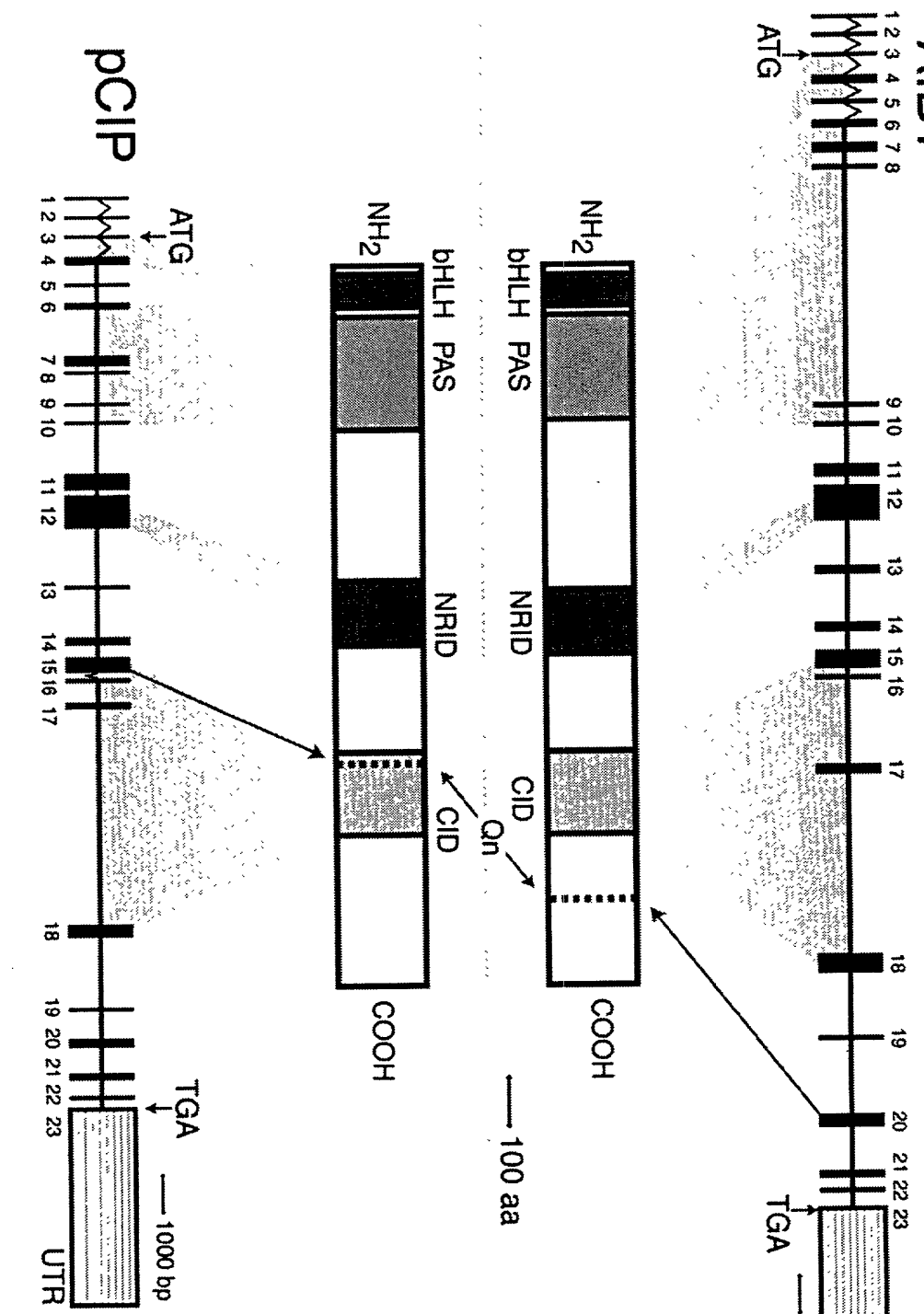
AIB1
ATG
TGA
1000 bp
UTR
NH2
bHLH
PAS
NRID
CID
Qn
COOH
100 aa
pCIP
ATG
TGA
1000 bp
UTR
NH2
bHLH
PAS
NRID
CID
COOH

FIGURE 5: MOUSE AIB1 (pCIP) INTRON/EXON BOUNDARIES

Exon	cDNA bp 5'exon	cDNA bp 3'exon	3'intron splice site	Exon sequence (5' to 3')	5'intron splice site
1		11		GGCGGCGAACG	
2	12			GATCAAAAGAATTTGCTGAA	
2		90		CCTTCCTGAACAGCTGTCAG	
3	91		TGTCACCTCTTCTTCCGCAG	TTGCTGATCTGTGATCAGGA	
3		195		TGTGATGCCCCAGGACAGGG	
4	196		GGCTTTTCTCCGCCTTCCAG	GCTTGTCTACAGTGGTGAGA	
4		368		ACGGCAAATAAAAGAACAAG	GTAACACAGAGTCAGAAAAA
5	369		GCTTCCTTCTGTGTCTTCAG	GAAAAACTATTTCCAGTGAT	
5		469		TAGGACCGCTTTTACTACAG	ATTTTCTTACAAACGAGGCT
6	470		ATTAACACATTCCACTGTAG	GCACTGGATGGTTTCCTGTT	
6		644		ACACTTACCAAAATCCACAG	GTGGGCTCTTCTTTGTGTTT
7	645		TTTTAATTTGTTTTTCAAAG	TTAATGGAGTTTCTTGGACT	
7		830		TATGCTGGAAGAAGGAGAAG	GTGAGAGGCGGGTCCACTGT
8	831		CTGGTGACCTTTCGTTGTAG	ACTTGCAGTGCTGTATGATC	
8		923		TACCAGACATGACCTTTCCG	GTAAGACCAGTCTTCACTGG
9	924		TCTGTTTTTATCTTTAATAG	GAAAGGTTGTCAATATAGAT	
9		1064		GAAGCGTCACTATCAAGAAG	GTGAGGGAGGCGTTTGGGGT
10	1065		GTGTGCTTCCCCCTCCGTAG	CTTATGTTCATGGCCACGCA	
10		1212		TCGACCCACTTTCTTCAGAG	GTGATGACACTAAAGCACCC
11	1213		TTGCGTGTGTTTGTTTGCAG	AGAACAGAATGGATACAGAC	
11		1589		CCAGTTCTCTCCTGCTGCAG	GTATCCACAGCTGCGTTTTC
12	1590		CGACCTTTCTCCATATGCAG	GTGCACACTCACCCATGGGA	
12		2458		AGACCGAGACGAACGAGGAG	GAGGTAAGGTACTCTCTGTT
13	2459		TTTAAAAGGTTCATTTTCAG	GTATCGGGAGACCTGGATAA	
13		2588		TGCAGGACCGAGTTCTCTGG	GTAAGGAAAACCAGAGTTTT
14	2589		AGCTTCTGTGTTTTCAACAG	GTTTGCGAAGTCCACAGCCT	
14		2783		GAATTACGGTGCCAACATGG	GTAGGTCATGTCTAAGTGTG

FIGURE 5: MOUSE AIB1 (pCIP) INTRON/EXON BOUNDARIES

Exon	cDNA bp 5'exon	cDNA bp 3'exon	3'intron splice site	Exon sequence (5' to 3')		5'intron splice site
15	2784		TGAGCCCTCCCTAATTTTAG	GCCCAAACAGAAATGTTCCT		
15		3095			GCAGCAGATGCTTCAAATGA	GTAAGCTGTCCCTTTCAATA
16	3096		ATTTTGATTTGCTCCCCCAG	GAACTGGTGAGATTCCCATG		
16		3222			CCTCACGGGTCTCAAAATAG	GTAGGGTTTTATTTTGGGAT
17	3223		TGACTCACGTCTCTCTCTAG	GCCTCTTCTTAGAAACTCTC		
17		3394			TTCCTGAGCTCGTGAATCAG	GTGGAGTTGCAATCTGTGAG
18	3395		CTTTGTGTTTGATGTTTAAG	GGACAAGCTTTGGAGTCCAA		
18		3688			AGAGGCTACAGGGCCAGCAG	GTAAGACCGGGCTGTCAGGG
19	3689		ACTAACCCAACTCTGTTCAG	TTTTTAAATCAGAGCCGGCA		
19		3772			TGAGGCCCATGATGCCCCAG	GTACGTTCCCTGCAGAGAAG
20	3773		TGTCTCTTGGCTACCAGCAG	GCTTTCTTTAATGCCCAAAT		
20		3989			TCCATATCCAGCAAATTACG	GTAAACCTGTCAGATTGTGC
21	3990		TTTCTGTTCATTTCTTTAAG	GAATGGGACAACCACCAGAG		
21		4164			GGGAACCTGGCCAGGAATGG	GTAAGGATGGGACTTACTTT
22	4165		CTGTTACCCTTTCTTTGCAG	CTCCTTCCCCCAGCAGCAGT		
22		4306			TGCCCATGGGCCCCGATCAG	GTACGGGCATCTATTCTTAC
23	4307		CTGTGTTCTTCTGTTAACAG	AAATACTGCTGACATCTCCC		
23		4622				

FIGURE 6: HUMAN AIB1 INTRON/EXON BOUNDARIES

Exon	cDNA bp 5'exon	cDNA bp 3'exon	3'intron splice site	Exon sequence (5' to 3')	5'intron splice site
1		102		GAGGAAAATGGCGGCGGGAG	GTGAGTGGAGATAAAGGAGG
2	103		CCTCTTCTTTTTGTCCTCAG	GATCAAAATACTTGCTGGAT	
2		181		TCCTTTGACTGGTTAGCCAG	GTAATTCAGCTTTAGTTTGA
3	182		TTCTCATTATTCTCTCTTAG	TTGCTGATGTATATTCAAGA	
3		283		TGTGATACTCCAGGACAAGG	GTAGGTGACTTATTTCCTGG
4	284		TTCTACGCCTTTTCCCTTAG	TCTTACCTGCAGTGGTGAAA	
4		456		ACGTCAAATAAAAGAGCAAG	GTAATAAAAACACTCATGTC
5	457		ACCACCTTCTGTCTTTTCAG	GAAAAACTATTTCCAATGAT	
5		557		TAGGACCGCTTTTACTTCAG	GCAAGTATAAAGATTTTAAC
6	558		ATTAACATATCCTATTTTAG	GCATTGGATGGTTTCCTATT	
6		732		GAATTTACCAAAATCTACAG	GTAGGCTTTTAATGTGTATT
7	733		TTTCAATTTGTTTTCCAAAG	TTAATGGAGTTTCCTGGACA	
7		921		TATGATGGAGGAAGGGGAAG	GTAAGAGCTATTATATGTTT
8	922		GGGTGAATTTTTTATTGTAG	ATTTGCAATCTTGTATGATC	
8		1023		TACCAGACATGATCTTTCAG	GTAAAAATCTTTTTTTGTCC
9	1024		TTCCTTTTTTTGTTTAATAG	GAAAGGTTGTCAATATAGAT	
9		1164		GAAACGTCACTATCAAGAAG	GTAAAGAATTTTGGGGTTGA
10	1165		TGGGATATTTTCCCCAACAG	CTTATCTTAATGGCCATGCA	
10		1312		TCAACCCACTTCCTTCAGAG	GTAATGATAGATTACTGTGT
11	1313		GTTTGATGTTTGTTTTGCAG	AGAACAGAATGGATATAGAC	
11		1704		TCAGTTTTCTCCTGTTGCAG	GTATTTGTGTTGACATTTCC
12	1705		AAATTTTTTTTCAAATTCAG	GTGTGCACTCTCCCATGGCA	
12		2576		AGACAGAGACAAGTGAAGAG	GTAATTTGTTTTCTGTATAT
13	2577		TTTTAAAACTTTATTTTCAG	GGATCTGGAGACTTGGATAA	
13		2712		TCAAGGAACTAATTCTCTGG	GTAAGAATGAACTAGGTTTT

FIGURE 6: HUMAN AIB1 INTRON/EXON BOUNDARIES

Exon	cDNA bp 5'exon	cDNA bp 3'exon	3'intron splice site	Exon sequence (5' to 3')	5'intron splice site
14	2713		TTGTATTGTGTTTTCAACAG	GTTTGAAAAGTTCACAGTCT	
14		2907			AAATTATGGCTCAAGTATGG GTATGTTATTTCTAATTAGT
15	2908		AGTATGGCTACCTGTTTTAG	GTGGGCCAAACCGAAATGTG	
15		3280			TCTCATGGCACTCAAAATAG GTGGGGTGTTATTTTGTGAC
16	3281		GATTGCAAGTCTTTTTCTAG	GCCTCTTCTTAGGAATTCCC	
16		3452			TTCCTGAACTTGTCAATCAG GTAGGTTGCATTAACATGGA
17	3453		TTTTATGTGTTGTGTTTAAG	GGACAGGCATTAGAGCCCAA	
17		3746			AGAGGCTGCAGGGCCAGCAG GTAACCAGTCATGTGTTCTT
18	3747		ACCAACTTGTCTCACCTCAG	TTTTTGAATCAGAGCCGACA	
18		3839			GGCCTATGATGCAGCCCCAG GTGAGCTCCCAGGTGAGGAT
19	3840		CACTCTTTCTTGGGTATTAG	CAGGGTTTTCTTAATGCTCA	
19		4134			TCCATATCAACCAAATTATG GTAAATCTGACAATGAAAAT
20	4135		TTCTGTTTTATTTTTGTAAG	GAATGGGACAACAACCAGAT	
20		4309			GGAAATTTGGCCAGGAACAG GTAAAGAACAGTGACTTATA
21	4310		TACCATTTGTTTACTTACAG	CTCCTTTTCCCAGCAGCAGT	
21		4450			TGCCTATGGGTCCTGATCAG GTATGGGATCGATTCCTTAC
22 (18)	4451		TTTTTCCTGGTTGCTGACAG	AAATACTGCTGACATCTCTG	

COMBINED DECLARATION AND POWER OF ATTORNEY FOR PATENT APPLICATION

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name,

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled AIB1, A NOVEL STEROID RECEPTOR CO-ACTIVATOR, the specification of which

[X] is attached hereto.

[ ] was filed on _____ as Application No. _____.

[ ] was described and claimed in PCT International Application No. _____, filed on _____, and as amended under PCT Article 19 on _____ (if applicable).

[ ] and was amended on _____ (if applicable).

[ ] with amendments through _____ (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, § 1.56. If this is a continuation-in-part application filed under the conditions specified in 35 U.S.C. § 120 which discloses and claims subject matter in addition to that disclosed in the prior copending application, I further acknowledge the duty to disclose material information as defined in 37 CFR § 1.56 which occurred between the filing date of the prior application and the national or PCT international filing date of the continuation-in-part application.

I hereby claim foreign priority benefits under Title 35, United States Code, § 119(a)-(d) of any foreign application(s) for patent or inventor's certificate or of an PCT International application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application(s) for patent or inventor's certificate or any PCT International application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) on which priority is claimed:

Prior Foreign Application(s)			Priority Claimed
US98/12689 (Number)	PCT (Country)	17/6/1998 (Day/Month/Year Filed)	[X] Yes [ ] No

I hereby claim the benefit under Title 35, United States Code, § 119(e) of any United States provisional application(s) listed below:

60/049,728 June 17, 1997
(Application No.) (Filing Date)

I hereby claim the benefit under Title 35, United States Code, § 120 of any United States application(s) or § 365(c) of any PCT International application(s) designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of Title 35, United States Code, § 112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, § 1.56(a) which occurred between the filing date of the prior application and the national or PCT International filing date of this application:

PCT/US98/12689 (Application No.) June 17, 1998 (Filing Date) pending (Status: patented, pending, abandoned)

The undersigned hereby authorizes the U.S. attorney or agent named herein to accept and follow instructions from _____ as to any action to be taken in the Patent and Trademark Office regarding this application without direct communication between the U.S. attorney or agent and the undersigned. In the event of a change in the persons from whom instructions may be taken, the U.S. attorney or agent named herein will be so notified by the undersigned.

I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application, to file a corresponding international application, and to transact all business in the Patent and Trademark Office connected therewith:

Name	Reg. No.	Name	Reg. No.
James C. Haight	25,588	Gloria Richmond	30,416
Robert Benson	33,612	Jack Spiegel	34,477
Susan S. Rucker	35,762	David R. Sadowski	32,808
John Peter Kim	38,514	Steve Ferguson	38,448
Stephen L. Finley	36,357		

with an Associate Power of Attorney to the following:

Name	Reg. No.	Name	Reg. No.
Richard J. Polley	28,107	Stacey C. Slater	36,011
William D. Noonan	30,878	Gregory V. Bean	36,448
William Y. Conwell	31,943	Joel R. Meyer	37,677
Mark L. Becker	31,325	David J. Earp	41,401
Donald L. Stephens Jr.	34,022		

all of the law firm of Klarquist Sparkman Campbell Leigh & Whinston, LLP.

Address all telephone calls to William D. Noonan, M.D., telephone number 503/226-7391 and facsimile number 503/228-9446.

Address all correspondence to:

KLARQUIST SPARKMAN CAMPBELL
LEIGH & WHINSTON, LLP
One World Trade Center, Suite 1600
121 SW Salmon Street
Portland, OR 97204-2988

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full name of sole or first inventor: Paul Meltzer

Inventor's signature Paul M [illegible] MD

Date 8/12/98

Residence: ~~Potomac~~ Rockville, Maryland

Citizenship: United States of America

Post Office address: ~~11702 Coldstream Drive, Potomac, Maryland 20854~~ 5906 Bloomingdale Terr. Rockville MD 20852

2-00

tFull name of second joint inventor, if any: Jeffrey Trent

Inventor's signature _____ 8/12/98 Date

Residence: Rockville, Maryland MD

Citizenship: United States of America

Post Office address: 10 Fairwood Court, Rockville, Maryland 20850